

HeshamMasterProjects

December 20, 2025

Mount Google Drive

```
[ ]: from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

DATASET URL <https://www.kaggle.com/datasets/sumaya23abdul/automobile-database>

Import libraries

```
[ ]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.impute import KNNImputer
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
import seaborn as sns

pd.set_option('display.max_columns', None)
pd.set_option('display.width', 220)
```

Explore Dataset

```
[ ]: import pandas as pd
df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳Automobile_data.csv')
df.head()
```

```
[ ]:      symboling normalized-losses      make fuel-type aspiration num-of-doors
body-style drive-wheels engine-location  wheel-base  length  width  height
curb-weight engine-type num-of-cylinders  engine-size  \
0          3          ?  alfa-romero      gas      std      two
convertible          rwd          front      88.6  168.8  64.1  48.8
2548          dohc          four      130
1          3          ?  alfa-romero      gas      std      two
convertible          rwd          front      88.6  168.8  64.1  48.8
2548          dohc          four      130
2          1          ?  alfa-romero      gas      std      two
hatchback          rwd          front      94.5  171.2  65.5  52.4
```

2823	ohcv	six	152				
3	2	164	audi	gas	std	four	
sedan	fwd	front	99.8	176.6	66.2	54.3	
2337	ohc	four	109				
4	2	164	audi	gas	std	four	
sedan	4wd	front	99.4	176.6	66.4	54.3	
2824	ohc	five	136				

		fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg
highway-mpg	price							
0	mpfi	3.47	2.68		9.0	111	5000	21
27	13495							
1	mpfi	3.47	2.68		9.0	111	5000	21
27	16500							
2	mpfi	2.68	3.47		9.0	154	5000	19
26	16500							
3	mpfi	3.19	3.4		10.0	102	5500	24
30	13950							
4	mpfi	3.19	3.4		8.0	115	5500	18
22	17450							

Explore Data Types

```
[ ]: df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳Automobile_data.csv')
df.dtypes
```

```
[ ]: symboling          int64
normalized-losses      object
make                   object
fuel-type              object
aspiration             object
num-of-doors           object
body-style             object
drive-wheels           object
engine-location        object
wheel-base            float64
length                float64
width                  float64
height                 float64
curb-weight            int64
engine-type            object
num-of-cylinders       object
engine-size            int64
fuel-system            object
bore                   object
stroke                 object
```

```

compression-ratio    float64
horsepower            object
peak-rpm              object
city-mpg              int64
highway-mpg          int64
price                 object
dtype: object

```

Explore Dataset Columns

```

[ ]: df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳Automobile_data.csv')
df.columns

[ ]: Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration',
'num-of-doors', 'body-style', 'drive-wheels', 'engine-location', 'wheel-base',
'length', 'width', 'height', 'curb-weight', 'engine-type',
'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg',
'price'],
          dtype='object')

```

check missing data

```

[ ]: df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳Automobile_data.csv')
df.isnull().sum()

[ ]: symboling          0
normalized-losses      0
make                   0
fuel-type              0
aspiration             0
num-of-doors           0
body-style             0
drive-wheels           0
engine-location        0
wheel-base            0
length                0
width                  0
height                 0
curb-weight            0
engine-type            0
num-of-cylinders       0
engine-size            0
fuel-system            0
bore                   0
stroke                 0

```

```

compression-ratio    0
horsepower            0
peak-rpm              0
city-mpg              0
highway-mpg          0
price                0
dtype: int64

```

handle missing data

```

[ ]: print("\nPercentage of missing data:")
      missing_percent = df.isnull().mean()*100
      print(missing_percent)

      threshold = 30
      cols_to_drop = missing_percent[missing_percent > threshold].index
      df.drop(columns=cols_to_drop)
      print("\nDropped columns (>30% missing):", cols_to_drop)
      print("Remaining columns:", df.columns.tolist())

```

Percentage of missing data:

```

symboling            0.0
normalized-losses    0.0
make                 0.0
fuel-type            0.0
aspiration            0.0
num-of-doors         0.0
body-style           0.0
drive-wheels         0.0
engine-location      0.0
wheel-base           0.0
length              0.0
width                0.0
height               0.0
curb-weight          0.0
engine-type          0.0
num-of-cylinders     0.0
engine-size          0.0
fuel-system          0.0
bore                 0.0
stroke              0.0
compression-ratio    0.0
horsepower           0.0
peak-rpm             0.0
city-mpg             0.0
highway-mpg          0.0
price               0.0

```

dtype: float64

Dropped columns (>30% missing): Index([], dtype='object')

Remaining columns: ['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration', 'num-of-doors', 'body-style', 'drive-wheels', 'engine-location', 'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type', 'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg', 'price']

Preprocessing

```
[ ]: label_encoder = LabelEncoder()
num_cols = df.select_dtypes(include=[np.number]).columns
for col in num_cols:
    df[col] = label_encoder.fit_transform(df[col])

cat_cols = df.select_dtypes(include=['object']).columns
for col in cat_cols:
    df[col] = label_encoder.fit_transform(df[col])
```

Filling missing values

```
[ ]: num_cols = df.select_dtypes(include=[np.number]).columns
for col in num_cols:
    mean_value = df[col].mean()
    df[col] = df[col].fillna(mean_value)

cat_cols = df.select_dtypes(include=['object']).columns
for col in cat_cols:
    mode_value = df[col].mode()[0]
    df[col] = df[col].fillna(mode_value)

print("\nMissing values after handling:")
print(df.isnull().sum())

clean_path = '/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳automobile_missing_data_handeled.csv'
df.to_csv(clean_path, index=False)
print(f"\n Cleaned dataset saved to: {clean_path}")
```

Missing values after handling:

symboling	0
normalized-losses	0
make	0
fuel-type	0

```

aspiration          0
num-of-doors        0
body-style          0
drive-wheels        0
engine-location     0
wheel-base         0
length             0
width              0
height             0
curb-weight         0
engine-type         0
num-of-cylinders    0
engine-size         0
fuel-system         0
bore               0
stroke             0
compression-ratio   0
horsepower          0
peak-rpm           0
city-mpg            0
highway-mpg         0
price              0
dtype: int64

```

Cleaned dataset saved to: /content/gdrive/MyDrive/Master-Project/ExtractedData/automobile_missing_data_handeled.csv

Handle Outliers

```

[ ]: # @title

df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳automobile_missing_data_handeled.csv')
# Select only numeric columns
num_cols = df.select_dtypes(include=[np.number]).columns.tolist()
print("\nNumeric columns for outlier detection:", num_cols)

# --- 1. IQR Method ---
def detect_outliers_iqr(data, col):
    Q1 = data[col].quantile(0.25)
    Q3 = data[col].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    return data[(data[col] < lower) | (data[col] > upper)]

# Detect and count outliers for each numeric column

```

```

print("\n--- IQR Method ---")
for col in num_cols:
    outliers = detect_outliers_iqr(df, col)
    print(f"{col}: {len(outliers)} outliers")

```

Numeric columns for outlier detection: ['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration', 'num-of-doors', 'body-style', 'drive-wheels', 'engine-location', 'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type', 'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg', 'price']

```

--- IQR Method ---
symboling: 0 outliers
normalized-losses: 0 outliers
make: 0 outliers
fuel-type: 0 outliers
aspiration: 0 outliers
num-of-doors: 0 outliers
body-style: 0 outliers
drive-wheels: 0 outliers
engine-location: 0 outliers
wheel-base: 0 outliers
length: 0 outliers
width: 0 outliers
height: 0 outliers
curb-weight: 0 outliers
engine-type: 0 outliers
num-of-cylinders: 0 outliers
engine-size: 0 outliers
fuel-system: 0 outliers
bore: 0 outliers
stroke: 0 outliers
compression-ratio: 0 outliers
horsepower: 0 outliers
peak-rpm: 0 outliers
city-mpg: 0 outliers
highway-mpg: 0 outliers
price: 0 outliers

```

```

[ ]: # --- 2. Z-score Method ---
def detect_outliers_zscore(data, col):
    mean = data[col].mean()
    std = data[col].std()
    z_score = (data[col] - mean) / std
    return data[(z_score > 3) | (z_score < -3)]

```

```

print("\n--- Z-score Outliers ---")
for col in num_cols:
    outliers = detect_outliers_zscore(df, col)
    print(f"{col}: {len(outliers)} outliers")

```

```

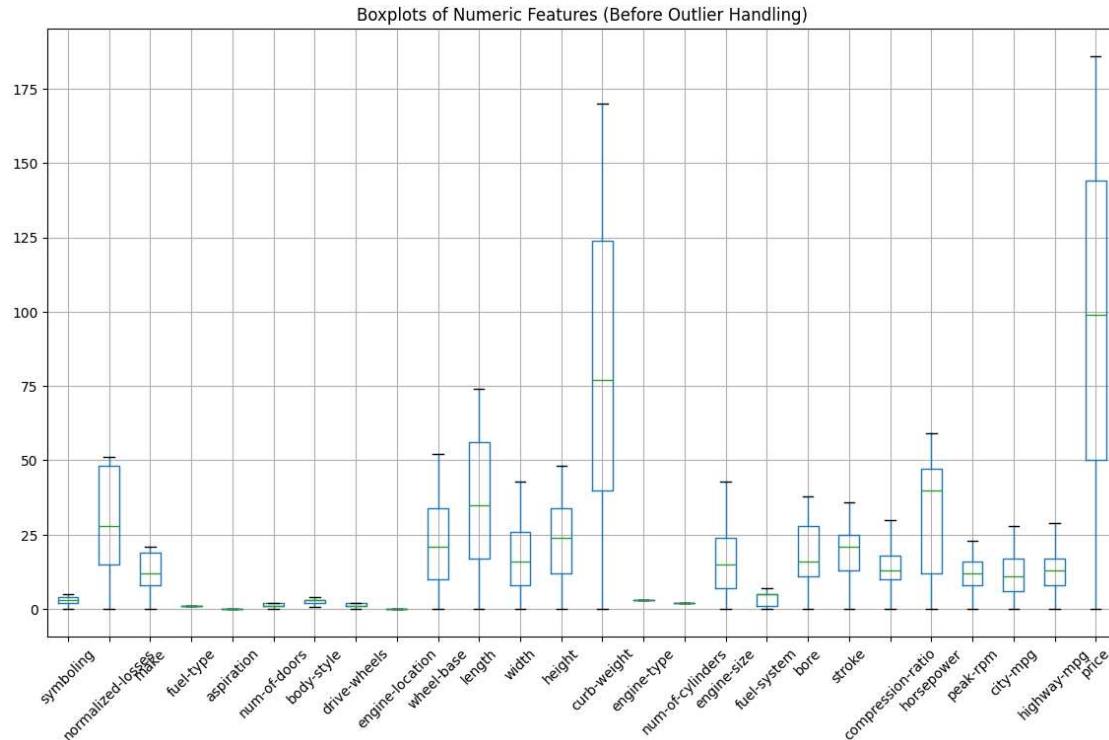
--- Z-score Outliers ---
symboling: 0 outliers
normalized-losses: 0 outliers
make: 0 outliers
fuel-type: 0 outliers
aspiration: 0 outliers
num-of-doors: 0 outliers
body-style: 0 outliers
drive-wheels: 0 outliers
engine-location: 0 outliers
wheel-base: 0 outliers
length: 0 outliers
width: 0 outliers
height: 0 outliers
curb-weight: 0 outliers
engine-type: 0 outliers
num-of-cylinders: 0 outliers
engine-size: 0 outliers
fuel-system: 0 outliers
bore: 0 outliers
stroke: 0 outliers
compression-ratio: 0 outliers
horsepower: 0 outliers
peak-rpm: 0 outliers
city-mpg: 0 outliers
highway-mpg: 0 outliers
price: 0 outliers

```

```

[ ]: # --- 3. Visualize with Boxplots ---
plt.figure(figsize=(14, 8))
df[num_cols].boxplot()
plt.title('Boxplots of Numeric Features (Before Outlier Handling)')
plt.xticks(rotation=45)
plt.show()

```



```
[ ]: # --- 4. Handle outliers by capping (winsorization) ---
for col in num_cols:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5 * IQR
    upper = Q3 + 1.5 * IQR
    df[col] = np.where(df[col] < lower, lower, np.where(df[col] > upper, upper,
    ↪df[col]))

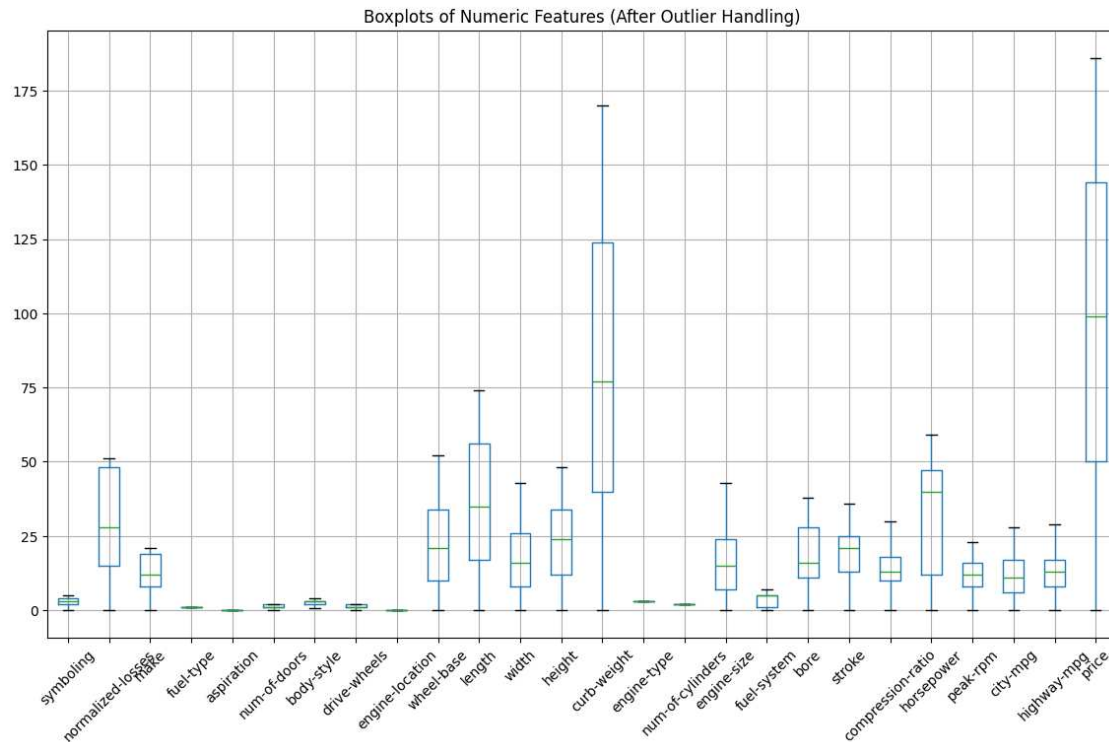
print("\nOutliers have been capped using IQR bounds.")

# --- 5. Verify after handling ---
plt.figure(figsize=(14, 8))
df[num_cols].boxplot()
plt.title('Boxplots of Numeric Features (After Outlier Handling)')
plt.xticks(rotation=45)
plt.show()

# --- Save cleaned dataset for visualization step ---
clean_outlier_path = '/content/gdrive/MyDrive/Master-Project/ExtractedData/
    ↪automobile_outliers_cleaned.csv'
df.to_csv(clean_outlier_path, index=False)
```

```
print(f"\n Outlier-handled dataset saved to: {clean_outlier_path}")
```

Outliers have been capped using IQR bounds.



Outlier-handled dataset saved to: /content/gdrive/MyDrive/Master-Project/ExtractedData/automobile_outliers_cleaned.csv

DATA VISUALIZATION

```
[ ]: # Load the outlier-handled dataset
df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳automobile_outliers_cleaned.csv')
print("Data shape:", df.shape)
df.head()

# Separate numeric and categorical columns
num_cols = df.select_dtypes(include=[np.number]).columns.tolist()
cat_cols = df.select_dtypes(include=['object']).columns.tolist()
print("\nNumeric Columns:", num_cols)
print("\nCategorical Columns:", cat_cols)
```

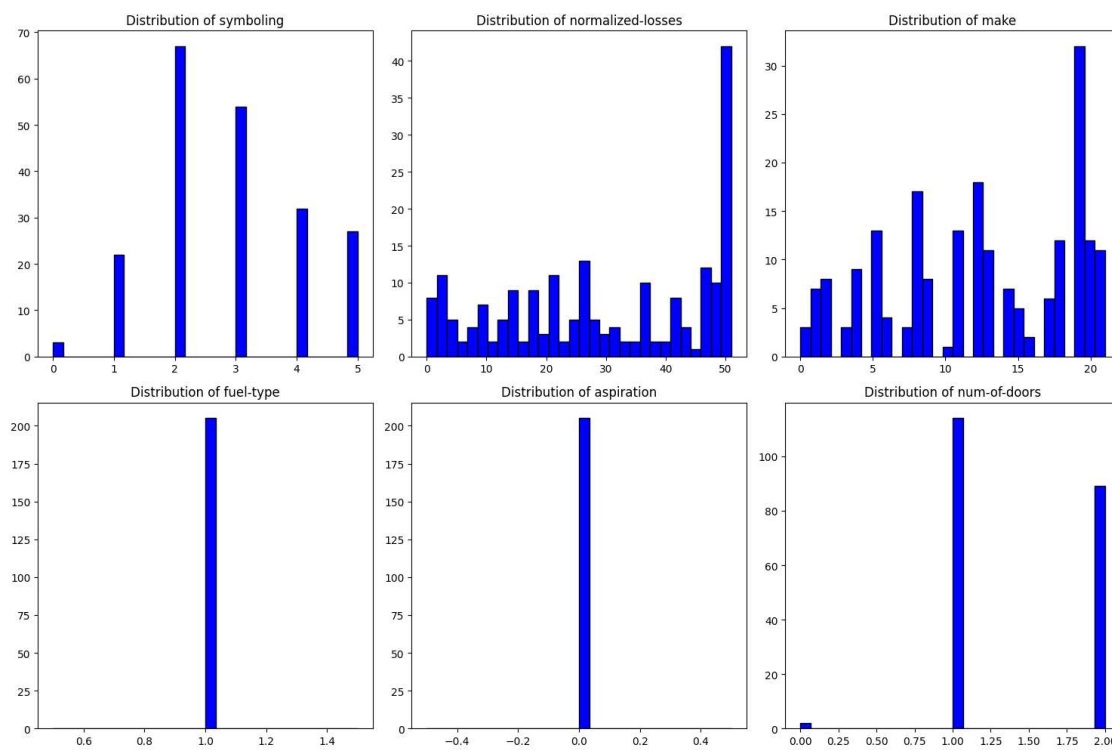
Data shape: (205, 26)

Numeric Columns: ['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration', 'num-of-doors', 'body-style', 'drive-wheels', 'engine-location', 'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type', 'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke', 'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg', 'highway-mpg', 'price']

Categorical Columns: []

```
[ ]: # 1. Histogram
```

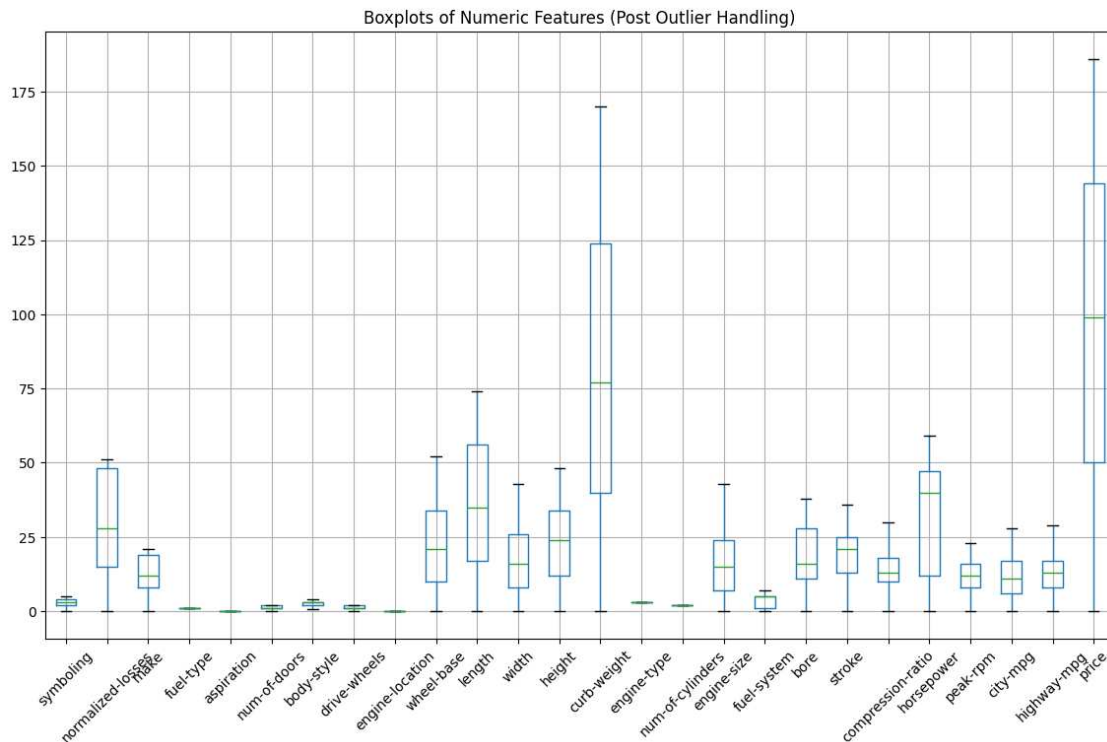
```
plt.figure(figsize=(15, 10))
for i, col in enumerate(num_cols[:6]): # first 6 columns
    plt.subplot(2, 3, i + 1)
    plt.hist(df[col], bins=30, color='blue', edgecolor='black')
    plt.title(f'Distribution of {col}')
plt.tight_layout()
plt.show()
```



```
[ ]: # 2. Boxplots
```

```
plt.figure(figsize=(14, 8))
df[num_cols].boxplot()
plt.title('Boxplots of Numeric Features (Post Outlier Handling)')
```

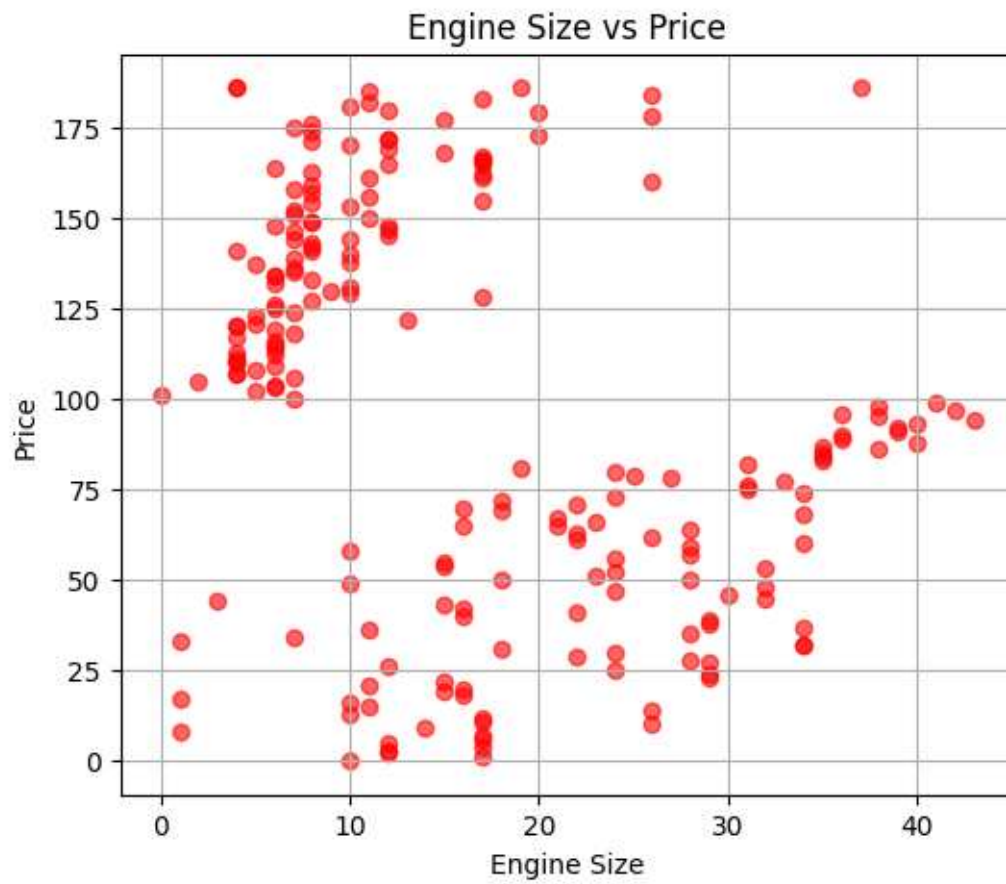
```
plt.xticks(rotation=45)
plt.show()
```

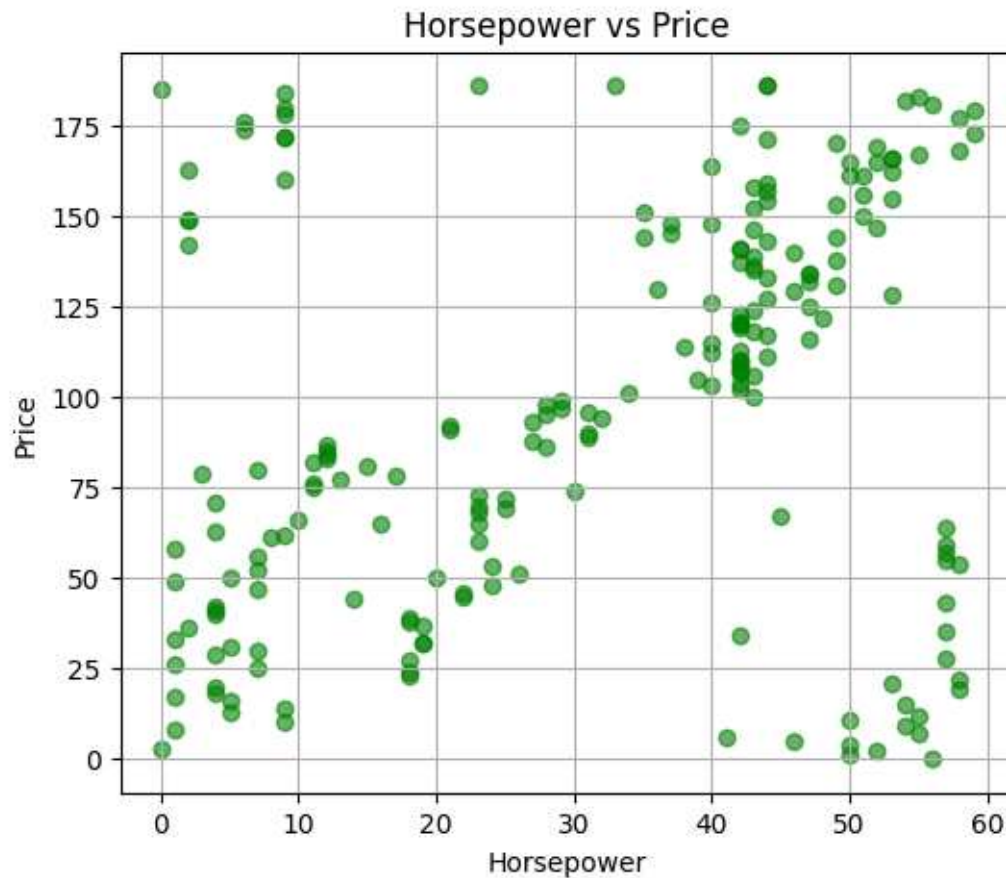


```
[ ]: # 3. Scatterplots
```

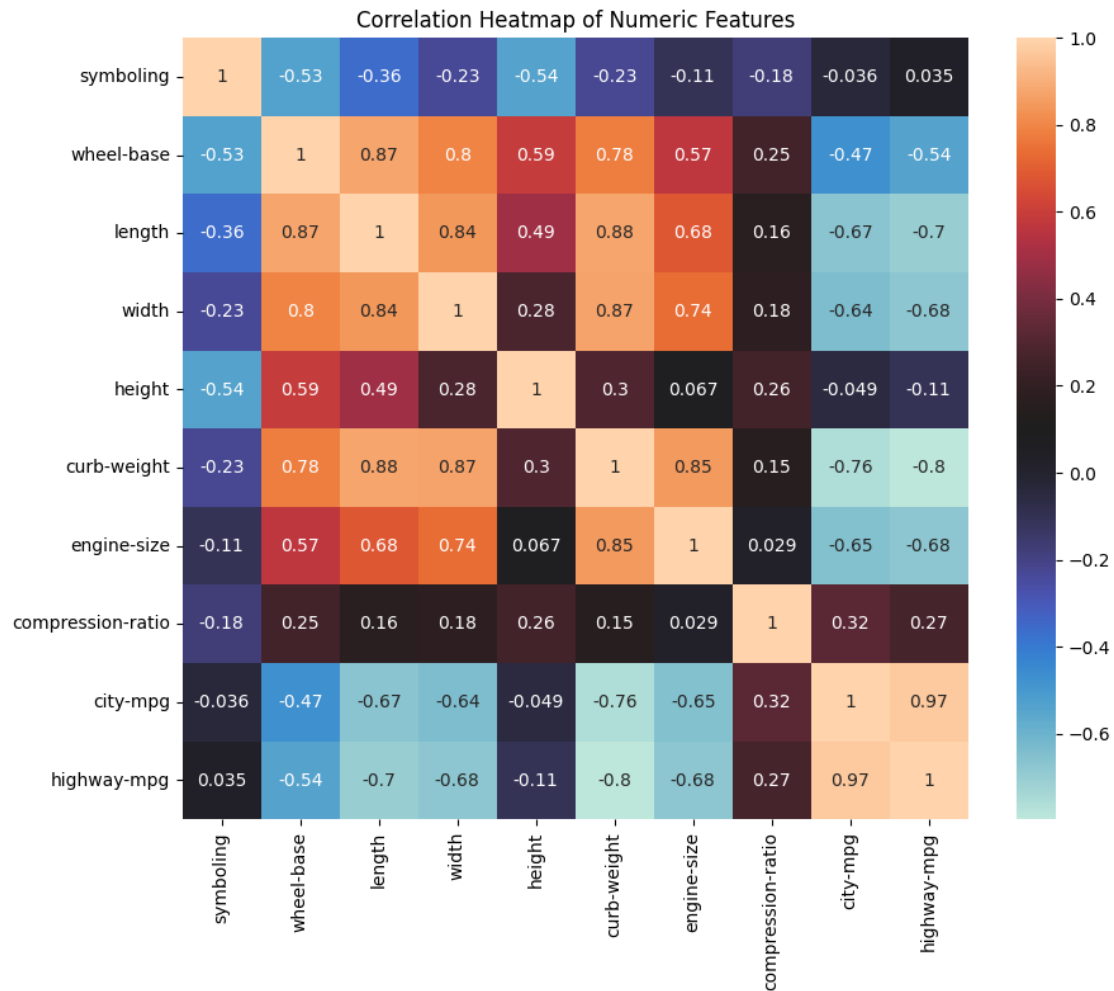
```
if 'engine-size' in df.columns and 'price' in df.columns:
    plt.figure(figsize=(6, 5))
    plt.scatter(df['engine-size'], df['price'], color='red', alpha=0.6)
    plt.title('Engine Size vs Price')
    plt.xlabel('Engine Size')
    plt.ylabel('Price')
    plt.grid(True)
    plt.show()

if 'horsepower' in df.columns and 'price' in df.columns:
    plt.figure(figsize=(6, 5))
    plt.scatter(df['horsepower'], df['price'], color='green', alpha=0.6)
    plt.title('Horsepower vs Price')
    plt.xlabel('Horsepower')
    plt.ylabel('Price')
    plt.grid(True)
    plt.show()
```





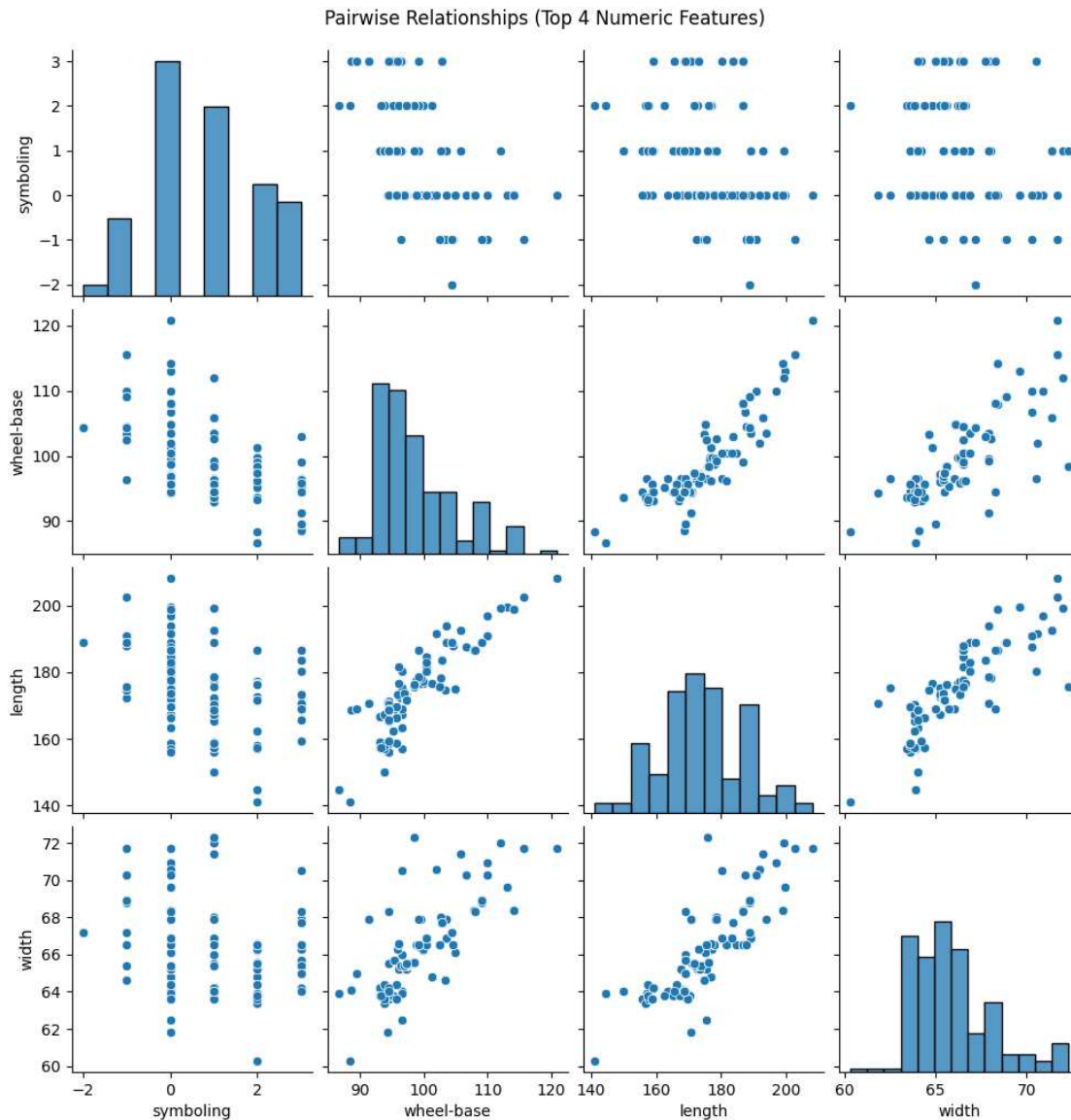
```
[ ]: # 4. Correlation Heatmap
df = pd.read_csv('/content/gdrive/MyDrive/Master-Project/ExtractedData/
↳Automobile_data.csv')
num_cols = df.select_dtypes([np.number]).columns.tolist()
plt.figure(figsize=(10,8))
corr_matrix = df[num_cols].corr()
sns.heatmap(corr_matrix, annot=True, cmap= 'icefire')
plt.title('Correlation Heatmap of Numeric Features')
plt.show()
```



```
[ ]: # 5. Pairplot (optional advanced)
print("\nPairplot")
plt.figure(figsize=(15, 9))
sns.pairplot(df[num_cols[:4]])
plt.suptitle('Pairwise Relationships (Top 4 Numeric Features)', y=1.02)
plt.show()
```

Pairplot

<Figure size 1500x900 with 0 Axes>



End-To-End Code

```
[3]: import os
import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder, StandardScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

```

from sklearn.linear_model import Ridge, Lasso
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.svm import SVR

from sklearn.feature_selection import SelectKBest, f_regression, SelectFromModel

import warnings
warnings.filterwarnings("ignore")

# ----- Save Path "change if required" -----
BASE_DIR = "/content/gdrive/MyDrive/Master-Project/ExtractedData/"
ORIGINAL_FILE = os.path.join(BASE_DIR, "Automobile_data.csv")

CLEANED_FILE = os.path.join(BASE_DIR, "automobile_missing_data_handled.csv")
RESULTS_ALL_FILE = os.path.join(BASE_DIR, "results_all_features.csv")
RESULTS_KBEST_FILE = os.path.join(BASE_DIR, "results_selectkbest.csv")
RESULTS_SFM_FILE = os.path.join(BASE_DIR, "results_selectfrommodel.csv")
RESULTS_SUMMARY_FILE = os.path.join(BASE_DIR, "results_summary.csv")
FEATURES_KBEST_FILE = os.path.join(BASE_DIR, "selected_features_kbest.csv")
FEATURES_SFM_FILE = os.path.join(BASE_DIR, "selected_features_selectfrommodel.
    ↪csv")

print("BASE_DIR:", BASE_DIR)
print("ORIGINAL_FILE:", ORIGINAL_FILE)

# ----- Load original dataset -----
df_raw = pd.read_csv(ORIGINAL_FILE)
print("Original shape:", df_raw.shape)

# -----normalize missing markers + numeric conversion where possible
df = df_raw.copy()
df.replace(["?", " ?", "NA", "N/A", "na", "n/a", ""], np.nan, inplace=True)

# Try converting object columns that are actually numeric
for col in df.columns:
    if df[col].dtype == "object":
        converted = pd.to_numeric(df[col], errors="ignore")
        # If conversion changes dtype to numeric, keep it
        if converted.dtype != "object":
            df[col] = converted

# ----- Choose target column -----
TARGET = "price"

# If target has missing, drop those rows
df = df.dropna(subset=[TARGET]).copy()

```

```

# If target is still object, try numeric
df[TARGET] = pd.to_numeric(df[TARGET], errors="coerce")
df = df.dropna(subset=[TARGET]).copy()

print("After dropping missing target:", df.shape)

# ----- Handle missing values (create CLEANED dataset and save it)
# -----
# IMPORTANT: For modelling we still use train-only imputers inside pipelines to
# ----- avoid leakage.

X_full = df.drop(columns=[TARGET])
y_full = df[TARGET]

numeric_cols = X_full.select_dtypes(include=["number"]).columns.tolist()
categorical_cols = [c for c in X_full.columns if c not in numeric_cols]

# Impute numeric with median, categorical with most_frequent, and keep
# ----- categorical as string
X_clean = X_full.copy()

if numeric_cols:
    num_imp = SimpleImputer(strategy="median")
    X_clean[numeric_cols] = num_imp.fit_transform(X_clean[numeric_cols])

if categorical_cols:
    cat_imp = SimpleImputer(strategy="most_frequent")
    X_clean[categorical_cols] = cat_imp.
    # ----- fit_transform(X_clean[categorical_cols]).astype(str)

# Reset index of X_clean to ensure proper concatenation with y_full.
# ----- reset_index()
X_clean = X_clean.reset_index(drop=True)
df_cleaned = pd.concat([X_clean, y_full.reset_index(drop=True)], axis=1)

df_cleaned.to_csv(CLEANED_FILE, index=False)
print(" Saved cleaned dataset to:", CLEANED_FILE)
print("Cleaned shape:", df_cleaned.shape)

# =====
# Helper: preprocessing for ML (NO leakage: impute inside CV pipeline)
# =====
from sklearn.base import clone

def make_preprocess(X):
    num_cols = X.select_dtypes(include=["number"]).columns.tolist()

```

```

cat_cols = [c for c in X.columns if c not in num_cols]

numeric_transformer = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="median")),
    ("scaler", StandardScaler()),
])

categorical_transformer = Pipeline(steps=[
    ("imputer", SimpleImputer(strategy="most_frequent")),
    ("onehot", OneHotEncoder(handle_unknown="ignore")),
])

pre = ColumnTransformer(
    transformers=[
        ("num", numeric_transformer, num_cols),
        ("cat", categorical_transformer, cat_cols),
    ],
    remainder="drop"
)

return pre, num_cols, cat_cols

def regression_metrics(y_true, y_pred):
    mse = mean_squared_error(y_true, y_pred)
    rmse = np.sqrt(mse)
    mae = mean_absolute_error(y_true, y_pred)
    r2 = r2_score(y_true, y_pred)
    return rmse, mae, r2

def run_gridsearch_models(X_train, y_train, X_test, y_test, models_and_grids,
    tag="ALL"):
    preprocess, _, _ = make_preprocess(X_train)
    rows = []
    best_estimators = {}

    for name, estimator, grid in models_and_grids:
        pipe = Pipeline(steps=[
            ("preprocess", preprocess),
            ("model", estimator),
        ])

        gs = GridSearchCV(
            pipe,
            param_grid=grid,
            scoring="neg_root_mean_squared_error",
            cv=5,
            n_jobs=-1
        )
        gs.fit(X_train, y_train)

```

```

best_estimators[name] = gs.best_estimator_
pred = gs.best_estimator_.predict(X_test)
rmse, mae, r2 = regression_metrics(y_test, pred)

rows.append({
    "Tag": tag,
    "Model": name,
    "RMSE": rmse,
    "MAE": mae,
    "R2": r2,
    "BestParams": gs.best_params_
})

print(f"[{tag}] {name} done | RMSE={rmse:.3f} | R2={r2:.3f}")

return pd.DataFrame(rows).sort_values("RMSE"), best_estimators

# =====
# ML models + GridSearch on CLEANED dataset
# =====
# ----- 0) Mount Drive -----
from google.colab import drive
drive.mount('/content/gdrive')

X = df_cleaned.drop(columns=[TARGET])
y = df_cleaned[TARGET]

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)

models_all = [
    ("Ridge", Ridge(random_state=42),
     {"model__alpha": [0.1, 1, 10, 50, 100]}),

    ("Lasso", Lasso(random_state=42, max_iter=20000),
     {"model__alpha": [0.001, 0.01, 0.1, 1, 5]}),

    ("RandomForest", RandomForestRegressor(random_state=42),
     {"model__n_estimators": [200, 400],
      "model__max_depth": [None, 10, 20],
      "model__min_samples_split": [2, 5],
      "model__min_samples_leaf": [1, 2]}),

    ("GradientBoosting", GradientBoostingRegressor(random_state=42),

```

```

        {"model__n_estimators": [200, 400],
         "model__learning_rate": [0.05, 0.1],
         "model__max_depth": [2, 3, 4],
         "model__subsample": [0.8, 1.0]}},

        ("SVR", SVR(),
         {"model__C": [1, 10, 50],
          "model__epsilon": [0.1, 0.2, 0.5],
          "model__gamma": ["scale", "auto"]}),
    ]

results_all, best_all = run_gridsearch_models(
    X_train, y_train, X_test, y_test, models_all, tag="ALL_FEATURES"
)

display(results_all)
results_all.to_csv(RESULTS_ALL_FILE, index=False)
print(" Saved:", RESULTS_ALL_FILE)

# =====
# Feature Selection Methods (2 methods)
# Method A: SelectKBest(f_regression) (filter)
# Method B: SelectFromModel(Lasso) (embedded)
# =====
def get_feature_names_after_preprocess(preprocess_fitted, num_cols, cat_cols):
    # numeric names (after scaler)
    num_names = preprocess_fitted.named_transformers_["num"].
    ↪get_feature_names_out(num_cols)
    # categorical names (after one-hot)
    ohe = preprocess_fitted.named_transformers_["cat"].named_steps["onehot"]
    cat_names = ohe.get_feature_names_out(cat_cols)
    return np.concatenate([num_names, cat_names])

def run_gridsearch_with_selector(X_train, y_train, X_test, y_test, selector,
    ↪models_and_grids, tag="FS"):
    preprocess, num_cols, cat_cols = make_preprocess(X_train)
    rows = []
    best_estimators = {}

    for name, estimator, grid in models_and_grids:
        pipe = Pipeline(steps=[
            ("preprocess", preprocess),
            ("select", selector),
            ("model", estimator),
        ])

```

```

gs = GridSearchCV(
    pipe,
    param_grid=grid,
    scoring="neg_root_mean_squared_error",
    cv=5,
    n_jobs=-1
)
gs.fit(X_train, y_train)

best_estimators[name] = gs.best_estimator_
pred = gs.best_estimator_.predict(X_test)
rmse, mae, r2 = regression_metrics(y_test, pred)

rows.append({
    "Tag": tag,
    "Model": name,
    "RMSE": rmse,
    "MAE": mae,
    "R2": r2,
    "BestParams": gs.best_params_
})

print(f"[{tag}] {name} done | RMSE={rmse:.3f} | R2={r2:.3f}")

return pd.DataFrame(rows).sort_values("RMSE"), best_estimators

# ---- Method A: SelectKBest (k is tuned inside GridSearch) ----
k_values = [10, 15, 20, 30, 40]

models_fs = [
    ("Ridge", Ridge(random_state=42),
     {"select__k": k_values,
      "model__alpha": [0.1, 1, 10, 100]}),

    ("RandomForest", RandomForestRegressor(random_state=42),
     {"select__k": k_values,
      "model__n_estimators": [200, 400],
      "model__max_depth": [None, 10, 20]}),

    ("GradientBoosting", GradientBoostingRegressor(random_state=42),
     {"select__k": k_values,
      "model__n_estimators": [200, 400],
      "model__learning_rate": [0.05, 0.1],
      "model__max_depth": [2, 3]}),
]

selector_kbest = SelectKBest(score_func=f_regression)

```

```

results_kbest, best_kbest = run_gridsearch_with_selector(
    X_train, y_train, X_test, y_test,
    selector=selector_kbest,
    models_and_grids=models_fs,
    tag="SELECT_KBEST"
)

display(results_kbest)
results_kbest.to_csv(RESULTS_KBEST_FILE, index=False)
print(" Saved:", RESULTS_KBEST_FILE)

# ---- Method B: SelectFromModel(Lasso) ----
selector_sfm = SelectFromModel(
    estimator=Lasso(random_state=42, max_iter=20000),
    threshold="median" # keep features above median importance
)

models_sfm = [
    ("Ridge", Ridge(random_state=42),
     {"select__estimator__alpha": [0.001, 0.01, 0.1, 1.0],
      "model__alpha": [0.1, 1, 10, 100]}),

    ("RandomForest", RandomForestRegressor(random_state=42),
     {"select__estimator__alpha": [0.001, 0.01, 0.1, 1.0],
      "model__n_estimators": [200, 400],
      "model__max_depth": [None, 10, 20]}),

    ("GradientBoosting", GradientBoostingRegressor(random_state=42),
     {"select__estimator__alpha": [0.001, 0.01, 0.1, 1.0],
      "model__n_estimators": [200, 400],
      "model__learning_rate": [0.05, 0.1],
      "model__max_depth": [2, 3]}),
]

results_sfm, best_sfm = run_gridsearch_with_selector(
    X_train, y_train, X_test, y_test,
    selector=selector_sfm,
    models_and_grids=models_sfm,
    tag="SELECT_FROM_MODEL_LASSO"
)

display(results_sfm)
results_sfm.to_csv(RESULTS_SFM_FILE, index=False)
print(" Saved:", RESULTS_SFM_FILE)

# =====

```

```

# Export "most important features" lists for both methods
# =====
def export_selected_features(best_pipeline, X_train):
    preprocess = best_pipeline.named_steps["preprocess"]
    selector = best_pipeline.named_steps["select"]

    # Fit preprocess alone to get feature names
    pre, num_cols, cat_cols = make_preprocess(X_train)
    pre_fitted = clone(pre).fit(X_train)

    all_names = get_feature_names_after_preprocess(pre_fitted, num_cols,
    ↪cat_cols)

    # fit selector on transformed training data (to align mask)
    Xt = pre_fitted.transform(X_train)
    selector_fitted = clone(selector).fit(Xt, y_train)

    mask = selector_fitted.get_support()
    selected_names = all_names[mask]
    return selected_names

# Best pipeline per FS approach (lowest RMSE)
best_kbest_model_name = results_kbest.iloc[0]["Model"]
best_sfm_model_name = results_sfm.iloc[0]["Model"]

best_kbest_pipe = best_kbest[best_kbest_model_name]
best_sfm_pipe = best_sfm[best_sfm_model_name]

selected_kbest = export_selected_features(best_kbest_pipe, X_train)
selected_sfm = export_selected_features(best_sfm_pipe, X_train)

pd.DataFrame({"feature": selected_kbest}).to_csv(FEATURES_KBEST_FILE,
    ↪index=False)
pd.DataFrame({"feature": selected_sfm}).to_csv(FEATURES_SFM_FILE, index=False)

print(" Saved selected features (KBest):", FEATURES_KBEST_FILE, "| count =",
    ↪len(selected_kbest))
print(" Saved selected features (SFM): ", FEATURES_SFM_FILE, "| count =",
    ↪len(selected_sfm))

# =====
# Overall summary: compare all approaches
# =====
summary = pd.concat([
    results_all.assign(Approach="All Features"),
    results_kbest.assign(Approach="SelectKBest"),
    results_sfm.assign(Approach="SelectFromModel(Lasso)")
])

```

```
], ignore_index=True).sort_values("RMSE")

display(summary)
summary.to_csv(RESULTS_SUMMARY_FILE, index=False)
print(" Saved summary:", RESULTS_SUMMARY_FILE)
```

```
BASE_DIR: /content/gdrive/MyDrive/Master-Project/ExtractedData/
ORIGINAL_FILE: /content/gdrive/MyDrive/Master-
Project/ExtractedData/Automobile_data.csv
Original shape: (205, 26)
Chosen TARGET = price
After dropping missing target: (201, 26)
  Saved cleaned dataset to: /content/gdrive/MyDrive/Master-
Project/ExtractedData/automobile_missing_data_handled.csv
Cleaned shape: (201, 26)
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).
[ALL_FEATURES] Ridge done | RMSE=3013.622 | R2=0.926
[ALL_FEATURES] Lasso done | RMSE=3202.293 | R2=0.916
[ALL_FEATURES] RandomForest done | RMSE=2986.884 | R2=0.927
[ALL_FEATURES] GradientBoosting done | RMSE=2316.563 | R2=0.956
[ALL_FEATURES] SVR done | RMSE=11616.036 | R2=-0.103
```

	Tag	Model	RMSE	MAE	R2 \
3	ALL_FEATURES	GradientBoosting	2316.563407	1516.929861	0.956137
2	ALL_FEATURES	RandomForest	2986.884109	1905.166162	0.927080
0	ALL_FEATURES	Ridge	3013.621803	1900.817503	0.925769
1	ALL_FEATURES	Lasso	3202.293001	2064.887625	0.916183
4	ALL_FEATURES	SVR	11616.035703	7385.410824	-0.102868

```
BestParams
3 {'model__learning_rate': 0.05, 'model__max_dep...
2 {'model__max_depth': None, 'model__min_samples...
0 {'model__alpha': 1}
1 {'model__alpha': 5}
4 {'model__C': 50, 'model__epsilon': 0.5, 'model...
```

```
Saved: /content/gdrive/MyDrive/Master-
Project/ExtractedData/results_all_features.csv
[SELECT_KBEST] Ridge done | RMSE=3292.717 | R2=0.911
[SELECT_KBEST] RandomForest done | RMSE=2941.971 | R2=0.929
[SELECT_KBEST] GradientBoosting done | RMSE=2595.964 | R2=0.945
```

	Tag	Model	RMSE	MAE	R2 \
2	SELECT_KBEST	GradientBoosting	2595.963713	1770.133013	0.944919
1	SELECT_KBEST	RandomForest	2941.971111	1855.314705	0.929257
0	SELECT_KBEST	Ridge	3292.716689	2191.598463	0.911383

```
BestParams
```

```

2 {'model__learning_rate': 0.05, 'model__max_dep...
1 {'model__max_depth': None, 'model__n_estimator...
0 {'model__alpha': 0.1, 'select__k': 40}

```

Saved: /content/gdrive/MyDrive/Master-
Project/ExtractedData/results_selectkbest.csv
[SELECT_FROM_MODEL_LASSO] Ridge done | RMSE=3231.239 | R2=0.915
[SELECT_FROM_MODEL_LASSO] RandomForest done | RMSE=2995.943 | R2=0.927
[SELECT_FROM_MODEL_LASSO] GradientBoosting done | RMSE=2354.307 | R2=0.955

	Tag	Model	RMSE	MAE	\
2	SELECT_FROM_MODEL_LASSO	GradientBoosting	2354.307186	1498.596402	
1	SELECT_FROM_MODEL_LASSO	RandomForest	2995.943289	1881.700681	
0	SELECT_FROM_MODEL_LASSO	Ridge	3231.238843	2052.218921	

	R2	BestParams
2	0.954696	{'model__learning_rate': 0.05, 'model__max_dep...
1	0.926637	{'model__max_depth': None, 'model__n_estimator...
0	0.914661	{'model__alpha': 1, 'select__estimator__alpha'...

Saved: /content/gdrive/MyDrive/Master-
Project/ExtractedData/results_selectfrommodel.csv
Saved selected features (KBest): /content/gdrive/MyDrive/Master-
Project/ExtractedData/selected_features_kbest.csv | count = 40
Saved selected features (SFM): /content/gdrive/MyDrive/Master-
Project/ExtractedData/selected_features_selectfrommodel.csv | count = 37

	Tag	Model	RMSE	MAE	\
0	ALL_FEATURES	GradientBoosting	2316.563407	1516.929861	
8	SELECT_FROM_MODEL_LASSO	GradientBoosting	2354.307186	1498.596402	
5	SELECT_KBEST	GradientBoosting	2595.963713	1770.133013	
6	SELECT_KBEST	RandomForest	2941.971111	1855.314705	
1	ALL_FEATURES	RandomForest	2986.884109	1905.166162	
9	SELECT_FROM_MODEL_LASSO	RandomForest	2995.943289	1881.700681	
2	ALL_FEATURES	Ridge	3013.621803	1900.817503	
3	ALL_FEATURES	Lasso	3202.293001	2064.887625	
10	SELECT_FROM_MODEL_LASSO	Ridge	3231.238843	2052.218921	
7	SELECT_KBEST	Ridge	3292.716689	2191.598463	
4	ALL_FEATURES	SVR	11616.035703	7385.410824	

	R2	BestParams	\
0	0.956137	{'model__learning_rate': 0.05, 'model__max_dep...	
8	0.954696	{'model__learning_rate': 0.05, 'model__max_dep...	
5	0.944919	{'model__learning_rate': 0.05, 'model__max_dep...	
6	0.929257	{'model__max_depth': None, 'model__n_estimator...	
1	0.927080	{'model__max_depth': None, 'model__min_samples...	
9	0.926637	{'model__max_depth': None, 'model__n_estimator...	
2	0.925769	{'model__alpha': 1}	
3	0.916183	{'model__alpha': 5}	
10	0.914661	{'model__alpha': 1, 'select__estimator__alpha'...	

```

7  0.911383          {'model__alpha': 0.1, 'select__k': 40}
4 -0.102868 {'model__C': 50, 'model__epsilon': 0.5, 'model...

```

```

          Approach
0          All Features
8  SelectFromModel(Lasso)
5          SelectKBest
6          SelectKBest
1          All Features
9  SelectFromModel(Lasso)
2          All Features
3          All Features
10 SelectFromModel(Lasso)
7          SelectKBest
4          All Features

```

Saved summary: /content/gdrive/MyDrive/Master-
Project/ExtractedData/results_summary.csv

Results & Discussion 1. Results

This section evaluates the performance of multiple machine learning models applied to the automobile price dataset after handling missing values. Models were assessed using RMSE, MAE, and R^2 , with GridSearchCV employed for hyperparameter optimization.

1.1 Performance on All Features

Among all evaluated models, Gradient Boosting achieved the best performance, with an R^2 of 0.956 and the lowest RMSE. This indicates a strong ability to capture nonlinear relationships and feature interactions present in the dataset. Random Forest also performed well but exhibited slightly higher error, likely due to higher variance compared to boosting.

Linear models (Ridge and Lasso) achieved reasonable performance, confirming that many relationships in the data are approximately linear. However, their performance was inferior to ensemble methods, highlighting the presence of nonlinear effects. Support Vector Regression (SVR) performed poorly, yielding a negative R^2 value, and was therefore excluded from further consideration.

1.2 Effect of Feature Selection

Two feature selection techniques were applied:

SelectKBest (filter-based)

SelectFromModel using Lasso (embedded)

SelectKBest reduced the feature space while maintaining strong performance, although a small decrease in accuracy was observed. In contrast, Lasso-based feature selection preserved nearly the same performance as the full-feature model, demonstrating its effectiveness in removing redundant features without significant loss of predictive power.

1.3 Important Features

Across both feature selection methods, the most influential predictors were consistently related to engine capacity and vehicle size, such as engine size, horsepower, curb weight, and width. Fuel

efficiency variables also showed importance but with an inverse relationship to price. These findings align with domain knowledge and confirm the validity of the modeling approach.

2. Discussion

The results demonstrate that ensemble learning methods outperform linear and kernel-based models for automobile price prediction. Gradient Boosting's superior performance can be attributed to its sequential learning strategy, which reduces bias and models complex patterns effectively.

Feature selection proved valuable not only for improving interpretability but also for reducing model complexity. Notably, Lasso-based feature selection achieved almost identical predictive performance to the full-feature Gradient Boosting model, indicating that a compact subset of features is sufficient for accurate price estimation.

The poor performance of SVR suggests that it is not well suited for this dataset, likely due to sensitivity to noise and limited scalability in high-dimensional, mixed-type feature spaces.

Overall, these results highlight a clear trade-off between maximum accuracy and model interpretability, with Gradient Boosting combined with Lasso-based feature selection offering the most balanced solution.

Install necessary packages to save Colab notebook to PDF

```
[4]: !apt-get install texlive-xetex texlive-fonts-recommended texlive-plain-generic
```

Reading package lists... Done

Building dependency tree... Done

Reading state information... Done

The following additional packages will be installed:

```
dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
libcommons-parent-java libfontbox-java libgs9 libgs9-common libidn12
libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1 libruby3.0
libsyntaxtex2 libteckit0 libtexlua53 libtexluajit2 libwoff1 libzip-0-13
lmodern poppler-data preview-latex-style rake ruby ruby-net-telnet
ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils
teckit tex-common tex-gyre texlive-base texlive-binaries texlive-latex-base
texlive-latex-extra texlive-latex-recommended texlive-pictures tipa
xfonts-encodings xfonts-utils
```

Suggested packages:

```
fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
| postscript-viewer perl-tk xpdf | pdf-viewer xzdec
texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
texlive-latex-extra-doc texlive-latex-recommended-doc texlive-luatex
texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
```

default-jre-headless tipa-doc

The following NEW packages will be installed:

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
libcommons-parent-java libfontbox-java libgs9 libgs9-common libidn12
libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1 libruby3.0
libsynctex2 libteckit0 libtexlua53 libtexluajit2 libwoff1 libzip-0-13
lmodern poppler-data preview-latex-style rake ruby ruby-net-telnet
ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration tiutils
teckit tex-common tex-gyre texlive-base texlive-binaries
texlive-fonts-recommended texlive-latex-base texlive-latex-extra
texlive-latex-recommended texlive-pictures texlive-plain-generic
texlive-xetex tipa xfonts-encodings xfonts-utils

0 upgraded, 53 newly installed, 0 to remove and 1 not upgraded.

Need to get 182 MB of archives.

After this operation, 571 MB of additional disk space will be used.

Get:1 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-droid-fallback all 1:6.0.1r16-1.1build1 [1,805 kB]

Get:2 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-lato all 2.0-2.1 [2,696 kB]

Get:3 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 poppler-data all 0.4.11-1 [2,171 kB]

Get:4 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 tex-common all 6.17 [33.7 kB]

Get:5 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-urw-base35 all 20200910-1 [6,367 kB]

Get:6 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9-common all 9.55.0~dfsg1-0ubuntu5.13 [753 kB]

Get:7 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libidn12 amd64 1.38-4ubuntu1 [60.0 kB]

Get:8 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libijs-0.35 amd64 0.35-15build2 [16.5 kB]

Get:9 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libjbig2dec0 amd64 0.19-3build2 [64.7 kB]

Get:10 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libgs9 amd64 9.55.0~dfsg1-0ubuntu5.13 [5,032 kB]

Get:11 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libkpathsea6 amd64 2021.20210626.59705-1ubuntu0.2 [60.4 kB]

Get:12 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 libwoff1 amd64 1.0.2-1build4 [45.2 kB]

Get:13 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 dvisvgm amd64 2.13.1-1 [1,221 kB]

Get:14 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-lmodern all 2.004.5-6.1 [4,532 kB]

Get:15 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 fonts-noto-mono all 20201225-1build1 [397 kB]

Get:16 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 fonts-texgyre all 20180621-3.1 [10.2 MB]

Get:17 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:18 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-parent-
java all 43-1 [10.8 kB]
Get:19 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libcommons-logging-
java all 1.2-2 [60.3 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libptexenc1
amd64 2021.20210626.59705-1ubuntu0.2 [39.1 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rubygems-integration
all 1.18 [5,336 B]
Get:22 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby3.0 amd64
3.0.2-7ubuntu2.11 [50.1 kB]
Get:23 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-rubygems
all 3.3.5-2ubuntu1.2 [228 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby amd64 1:3.0~exp1
[5,100 B]
Get:25 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 rake all 13.0.6-2 [61.7
kB]
Get:26 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:27 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-webrick
all 1.7.0-3ubuntu0.2 [52.5 kB]
Get:28 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 ruby-xmlrpc all
0.3.2-1ubuntu0.1 [24.9 kB]
Get:29 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libruby3.0
amd64 3.0.2-7ubuntu2.11 [5,114 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libsinctex2
amd64 2021.20210626.59705-1ubuntu0.2 [55.6 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libteckit0 amd64
2.5.11+ds1-1 [421 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexlua53
amd64 2021.20210626.59705-1ubuntu0.2 [120 kB]
Get:33 <http://archive.ubuntu.com/ubuntu> jammy-updates/main amd64 libtexluajit2
amd64 2021.20210626.59705-1ubuntu0.2 [267 kB]
Get:34 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 libzip-0-13 amd64
0.13.72+dfsg.1-1.1 [27.0 kB]
Get:35 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-encodings all
1:1.0.5-0ubuntu2 [578 kB]
Get:36 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 xfonts-utils amd64
1:7.7+6build2 [94.6 kB]
Get:37 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 lmodern all
2.004.5-6.1 [9,471 kB]
Get:38 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 preview-latex-style
all 12.2-1ubuntu1 [185 kB]
Get:39 <http://archive.ubuntu.com/ubuntu> jammy/main amd64 t1utils amd64
1.41-4build2 [61.3 kB]
Get:40 <http://archive.ubuntu.com/ubuntu> jammy/universe amd64 teckit amd64
2.5.11+ds1-1 [699 kB]

```

Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all
20180621-3.1 [6,209 kB]
Get:42 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-
binaries amd64 2021.20210626.59705-1ubuntu0.2 [9,860 kB]
Get:43 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all
2021.20220204-1 [21.0 MB]
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-
recommended all 2021.20220204-1 [4,972 kB]
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base
all 2021.20220204-1 [1,128 kB]
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all
1:1.8.16-2 [207 kB]
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all
1:1.8.16-2 [5,199 kB]
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-
recommended all 2021.20220204-1 [14.4 MB]
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures
all 2021.20220204-1 [8,720 kB]
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra
all 2021.20220204-1 [13.9 MB]
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-
generic all 2021.20220204-1 [27.5 MB]
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21
[2,967 kB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all
2021.20220204-1 [12.4 MB]
Fetched 182 MB in 8s (23.6 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages ...
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 117528 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb
...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
Unpacking fonts-lato (2.0-2.1) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
Unpacking poppler-data (0.4.11-1) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.17_all.deb ...
Unpacking tex-common (6.17) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
Unpacking fonts-urw-base35 (20200910-1) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.13_all.deb ...

```

```

Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.13) ...
Selecting previously unselected package libidn12:amd64.
Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15build2) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.13_amd64.deb ...
Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.13) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build4) ...
Selecting previously unselected package dvisvgm.
Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
Unpacking dvisvgm (2.13.1-1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
Unpacking fonts-lmodern (2.004.5-6.1) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
Unpacking fonts-noto-mono (20201225-1build1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../17-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../18-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../19-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../20-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...

```

```

Selecting previously unselected package ruby3.0.
Preparing to unpack .../21-ruby3.0_3.0.2-7ubuntu2.11_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.11) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../22-ruby-rubygems_3.3.5-2ubuntu1.2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2ubuntu1.2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../23-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../24-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../25-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../26-ruby-webrick_1.7.0-3ubuntu0.2_all.deb ...
Unpacking ruby-webrick (1.7.0-3ubuntu0.2) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../27-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../28-libruby3.0_3.0.2-7ubuntu2.11_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.11) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../29-libsyntax2_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libsyntax2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../30-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../31-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb
...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack
.../32-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../33-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../34-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../35-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...

```

Selecting previously unselected package lmodern.
 Preparing to unpack .../36-lmodern_2.004.5-6.1_all.deb ...
 Unpacking lmodern (2.004.5-6.1) ...
 Selecting previously unselected package preview-latex-style.
 Preparing to unpack .../37-preview-latex-style_12.2-1ubuntu1_all.deb ...
 Unpacking preview-latex-style (12.2-1ubuntu1) ...
 Selecting previously unselected package t1utils.
 Preparing to unpack .../38-t1utils_1.41-4build2_amd64.deb ...
 Unpacking t1utils (1.41-4build2) ...
 Selecting previously unselected package teckit.
 Preparing to unpack .../39-teckit_2.5.11+ds1-1_amd64.deb ...
 Unpacking teckit (2.5.11+ds1-1) ...
 Selecting previously unselected package tex-gyre.
 Preparing to unpack .../40-tex-gyre_20180621-3.1_all.deb ...
 Unpacking tex-gyre (20180621-3.1) ...
 Selecting previously unselected package texlive-binaries.
 Preparing to unpack .../41-texlive-binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
 Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
 Selecting previously unselected package texlive-base.
 Preparing to unpack .../42-texlive-base_2021.20220204-1_all.deb ...
 Unpacking texlive-base (2021.20220204-1) ...
 Selecting previously unselected package texlive-fonts-recommended.
 Preparing to unpack .../43-texlive-fonts-recommended_2021.20220204-1_all.deb ...
 Unpacking texlive-fonts-recommended (2021.20220204-1) ...
 Selecting previously unselected package texlive-latex-base.
 Preparing to unpack .../44-texlive-latex-base_2021.20220204-1_all.deb ...
 Unpacking texlive-latex-base (2021.20220204-1) ...
 Selecting previously unselected package libfontbox-java.
 Preparing to unpack .../45-libfontbox-java_1%3a1.8.16-2_all.deb ...
 Unpacking libfontbox-java (1:1.8.16-2) ...
 Selecting previously unselected package libpdfbox-java.
 Preparing to unpack .../46-libpdfbox-java_1%3a1.8.16-2_all.deb ...
 Unpacking libpdfbox-java (1:1.8.16-2) ...
 Selecting previously unselected package texlive-latex-recommended.
 Preparing to unpack .../47-texlive-latex-recommended_2021.20220204-1_all.deb ...
 Unpacking texlive-latex-recommended (2021.20220204-1) ...
 Selecting previously unselected package texlive-pictures.
 Preparing to unpack .../48-texlive-pictures_2021.20220204-1_all.deb ...
 Unpacking texlive-pictures (2021.20220204-1) ...
 Selecting previously unselected package texlive-latex-extra.
 Preparing to unpack .../49-texlive-latex-extra_2021.20220204-1_all.deb ...
 Unpacking texlive-latex-extra (2021.20220204-1) ...
 Selecting previously unselected package texlive-plain-generic.
 Preparing to unpack .../50-texlive-plain-generic_2021.20220204-1_all.deb ...
 Unpacking texlive-plain-generic (2021.20220204-1) ...
 Selecting previously unselected package tipa.
 Preparing to unpack .../51-tipa_2%3a1.3-21_all.deb ...

```

Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../52-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.18) ...
Setting up libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up ruby-webrick (1.7.0-3ubuntu0.2) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.13) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.13) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...

```

```

/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-
ini-files/pdftexconfig.tex
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...
Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...
Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.11) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.11) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2ubuntu1.2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for mailcap (3.70+nmu1ubuntu1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.11) ...
/sbin/ldconfig.real: /usr/local/lib/libhwloc.so.15 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_opencl.so.0 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtcm.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic
link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

```

```

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero_v2.so.0 is not a
symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm_debug.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libumf.so.1 is not a symbolic link

Processing triggers for tex-common (6.17) ...
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time... done.

```

Change the Path to Notebook location

Save to PDF

```
[ ]: !jupyter nbconvert --to pdf HeshamMasterProjects.ipynb
```

```

[NbConvertApp] Converting notebook HeshamMasterProjects.ipynb to pdf
[NbConvertApp] Support files will be in HeshamMasterProjects_files/
[NbConvertApp] Making directory ./HeshamMasterProjects_files
[NbConvertApp] Writing 58862 bytes to notebook.tex
[NbConvertApp] Building PDF
[NbConvertApp] Running xelatex 3 times: ['xelatex', 'notebook.tex', '-quiet']
[NbConvertApp] Running bibtex 1 time: ['bibtex', 'notebook']
[NbConvertApp] WARNING | bibtex had problems, most likely because there were no
citations
[NbConvertApp] PDF successfully created
[NbConvertApp] Writing 559173 bytes to HeshamMasterProjects.pdf

```