

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv('global_llm_benchmark_dataset_models.csv')
```

```
print(df.isnull().sum())
```

```
model_name          0
provider            0
release_date        0
context_window      0
api_input_cost_per_1k_usd  0
api_output_cost_per_1k_usd 0
open_source         0
mmlu_score          0
gsm8k_score         0
humaneval_score     0
dtype: int64
```

```
df['api_input_cost_per_1k_usd'] = df['api_input_cost_per_1k_usd'].fillna(0)
df['api_output_cost_per_1k_usd'] = df['api_output_cost_per_1k_usd'].fillna(0)
df['mmlu_score'] = df['mmlu_score'].fillna(df['mmlu_score'].mean())
```

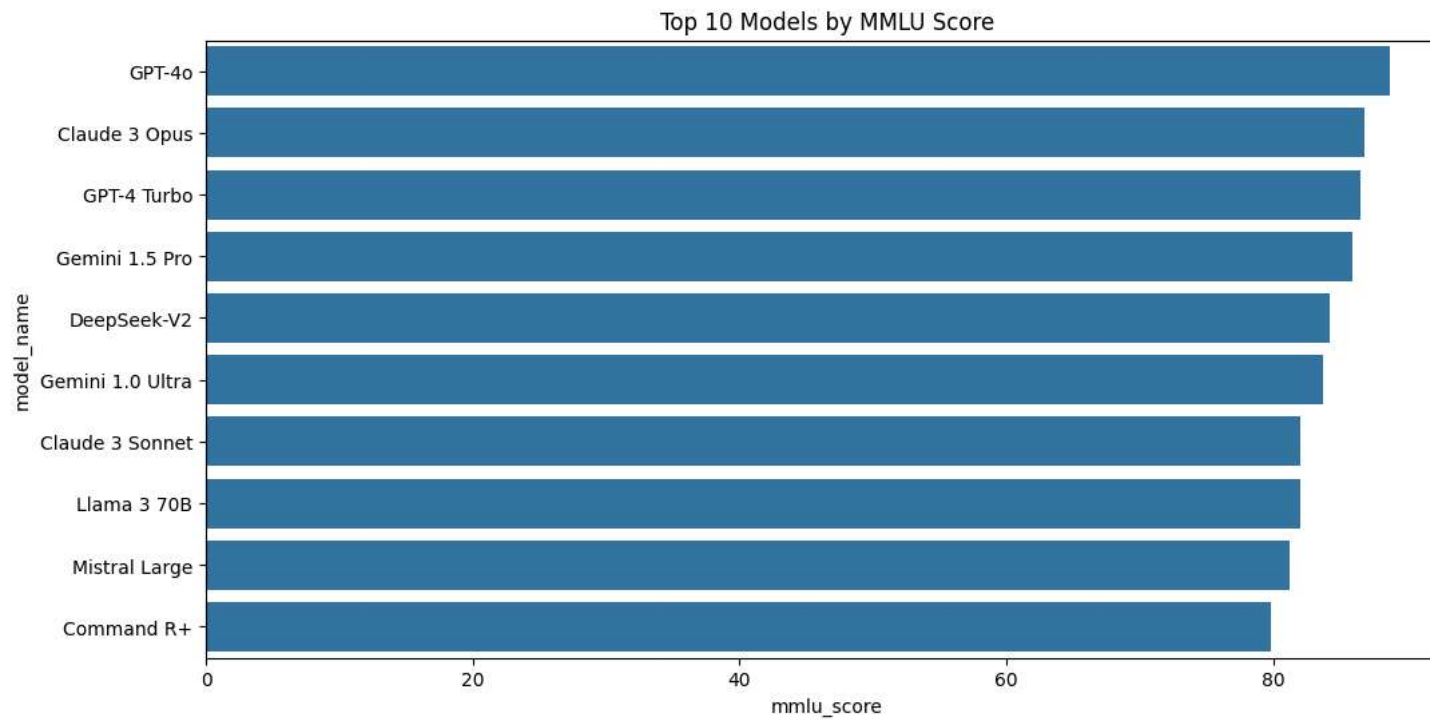
```
df.dropna(subset=['model_name'], inplace=True)
```

```
Q1 = df['context_window'].quantile(0.25)
Q3 = df['context_window'].quantile(0.75)
IQR = Q3 - Q1
```

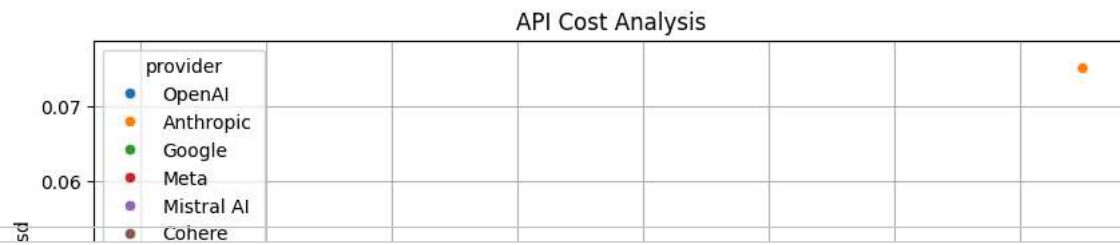
```
lower_limit = Q1 - 1.5 * IQR
upper_limit = Q3 + 1.5 * IQR
```

```
df_cleaned = df[(df['context_window'] >= lower_limit) & (df['context_window'] <= upper_limit)]
```

```
plt.figure(figsize=(12, 6))
sns.barplot(x='mmlu_score', y='model_name', data=df.sort_values('mmlu_score', ascending=False).head(10))
plt.title('Top 10 Models by MMLU Score')
plt.show()
```



```
plt.figure(figsize=(10, 6))
sns.scatterplot(x='api_input_cost_per_1k_usd', y='api_output_cost_per_1k_usd', hue='provider', data=df)
plt.title('API Cost Analysis')
plt.grid(True)
plt.show()
```



```
plt.figure(figsize=(8, 5))
sns.boxplot(x=df['context_window'])
plt.title('Context Window Distribution')
plt.show()
```

