1. **Bar Chart: Model Performance by Evaluation Metric**

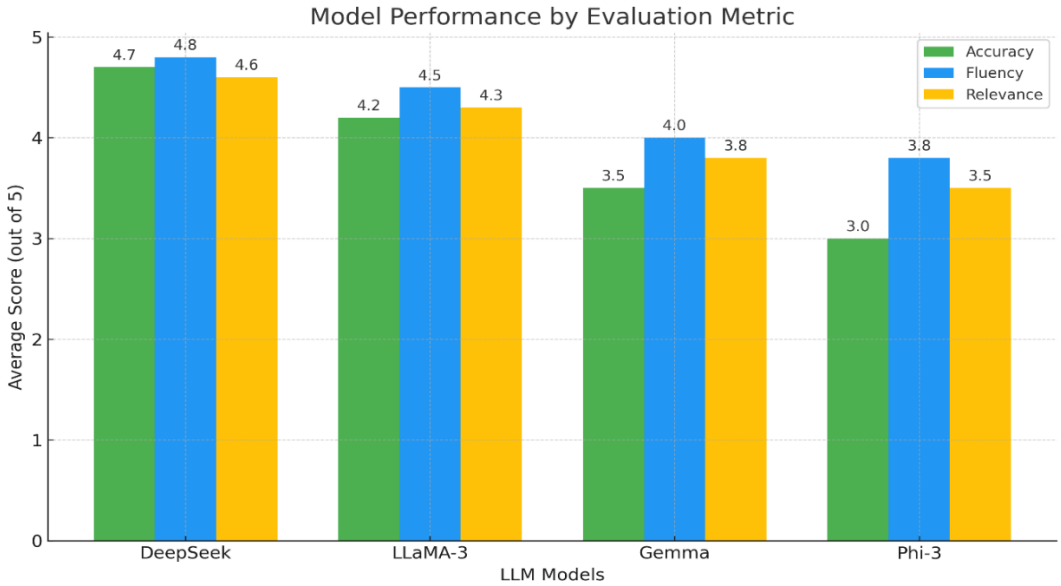
```
import matplotlib.pyplot as plt
import numpy as np

models = ['DeepSeek', 'LLaMA-3', 'Gemma', 'Phi-3']
accuracy = [4.7, 4.2, 3.5, 3.0]
fluency = [4.8, 4.5, 4.0, 3.8]
relevance = [4.6, 4.3, 3.8, 3.5]

      x = np.arange(len(models))
      width = 0.25

      plt.figure(figsize=(10, 6))
      plt.bar(x - width, accuracy, width, label='Accuracy',
color='mediumseagreen')
      plt.bar(x, fluency, width, label='Fluency', color='dodgerblue')
      plt.bar(x + width, relevance, width, label='Relevance',
color='gold')

      plt.ylabel('Average Score (out of 5)')
      plt.xlabel('LLM Models')
      plt.title('Model Performance by Evaluation Metric')
      plt.xticks(x, models)
      plt.ylim(0, 5)
      plt.legend()
      plt.grid(axis='y', linestyle='--', alpha=0.7)
      plt.tight_layout()
      plt.show()
```

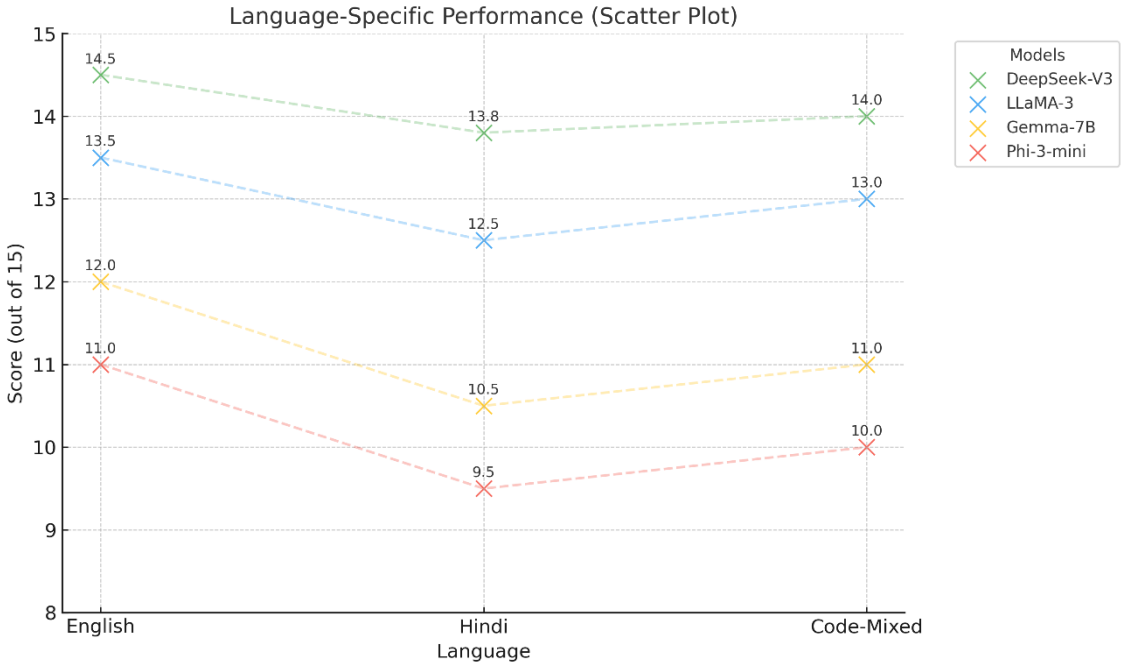


2. **Scatter Plot: Language-Specific Performance**

```
languages = ['English', 'Hindi', 'Code-Mixed']
deepseek = [14.5, 13.8, 14.0]
llama = [13.5, 12.5, 13.0]
gemma = [12.0, 10.5, 11.0]
phi = [11.0, 9.5, 10.0]
```

```
plt.figure(figsize=(10, 6))
plt.plot(languages, deepseek, marker='x', label='DeepSeek-V3', linestyle='-
-', color='green')
plt.plot(languages, llama, marker='x', label='LLaMA-3', linestyle='--',
color='dodgerblue')
plt.plot(languages, gemma, marker='x', label='Gemma-7B', linestyle='--',
color='gold')
plt.plot(languages, phi, marker='x', label='Phi-3-mini', linestyle='--',
color='red')

plt.title('Language-Specific Performance (Scatter Plot)')
plt.ylabel('Score (out of 15)')
plt.ylim(8, 15)
plt.legend(title='Models')
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()
```



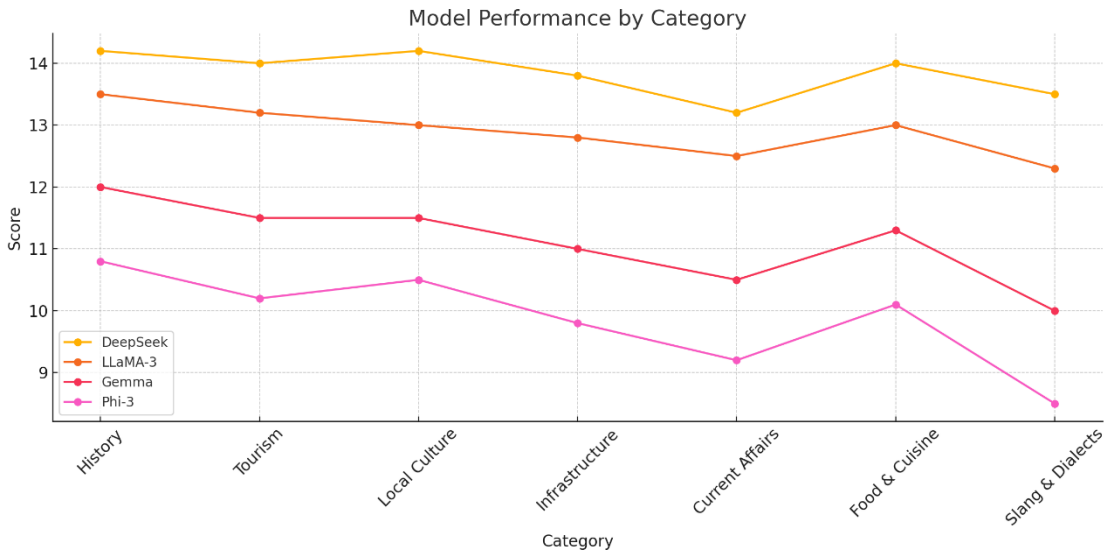
3. **Line Chart: Model Performance by Category**

```
categories = ['History', 'Tourism', 'Local Culture', 'Infrastructure',
'Current Affairs', 'Food & Cuisine', 'Slang & Dialects']
deepseek = [14.2, 14.0, 14.3, 13.8, 13.2, 14.0, 13.5]
llama = [13.5, 13.2, 13.0, 12.8, 12.5, 13.0, 12.3]
gemma = [12.0, 11.5, 11.5, 11.0, 10.5, 11.3, 10.0]
phi = [10.8, 10.2, 10.5, 9.8, 9.2, 10.1, 8.5]

plt.figure(figsize=(12, 6))
plt.plot(categories, deepseek, marker='o', label='DeepSeek',
color='orange')
plt.plot(categories, llama, marker='o', label='LLaMA-3', color='orangered')
plt.plot(categories, gemma, marker='o', label='Gemma', color='brown')
plt.plot(categories, phi, marker='o', label='Phi-3', color='hotpink')

plt.title('Model Performance by Category')
plt.ylabel('Score')
plt.xticks(rotation=30)
```

```
plt.legend()
plt.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
plt.show()
```



4. Error Pattern Analysis

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

# Data preparation
models = ['LLAMA', 'Genma', 'Phi']
error_types = ['Factual Errors', 'Linguistic Errors', 'Contextual Errors',
'Other Errors']

# The data matrix from your example (transposed to match the description)
data = np.array([
    [10, 15, 22, 23],   # Column 5
    [9, 18, 24, 25],     # Column 6
    [11, 16, 22, 26],    # Column 7
    [12.5, 17.5, 20.0, 21.0]  # Column (unnamed)
]).T

# Create the heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(data,
            annot=True,
            fmt=".1f",
            cmap="YlOrRd",  # Color tree (yellow-orange-red)
            xticklabels=['5', '6', '7', ''],  # Column labels
            yticklabels=error_types)  # Row labels

# Add titles and labels
plt.title('Heatmap of Error Distribution by Model and Type', pad=20)
plt.xlabel('Model')
plt.ylabel('Error Type')
```

```python
# Add model names on top
for i, model in enumerate(models):
    plt.text(i + 0.5, -0.5, model,
             ha='center', va='center',
             fontweight='bold')

# Adjust layout to prevent cutoff
plt.tight_layout()
plt.show()
```