

Practical 8: Data Visualization I - Cheatsheet

Theory & Libraries

Data Visualization: Graphical representation of data to uncover patterns, trends and outliers.

Seaborn: High-level interface for drawing attractive statistical graphics on top of Matplotlib.

Key Plots:

- countplot: Categorical counts
- histplot/displot: Distribution of numerical data
- boxplot/violinplot: Distribution with quartiles/outliers
- pairplot: Pairwise relationships
- jointplot: Bivariate analysis with marginal distributions

Dataset: Titanic (891 rows)

Columns: Survived, Pclass, Sex, Age, SibSp, Parch, Fare, Embarked, etc.

Code with Comments

```
# Import libraries
import pandas as pd          # Data manipulation
import numpy as np          # Numerical operations
import seaborn as sns        # Statistical plotting
import matplotlib.pyplot as plt # Plot display

# Load Titanic dataset
df = sns.load_dataset('titanic') # Inbuilt seaborn dataset
df.head()                        # Display first rows

# Check missing values and data types
df.info()                       # Info on nulls and types
df.describe()                   # Summary statistics

# 1. Survival count
sns.countplot(data=df, x='survived') # Count of survivors vs non-survivors
plt.title("Survival Count")
plt.show()

# 2. Passenger class vs gender
sns.countplot(data=df, x='pclass', hue='sex')
plt.title("Class by Gender")
plt.show()

# 3. Age distribution by survival
sns.boxplot(data=df, x='survived', y='age')
plt.title("Age vs Survival")
plt.show()

# 4. Fare distribution by class
sns.violinplot(data=df, x='pclass', y='fare')
plt.title("Fare by Class")
```

```
plt.show()
```

```
# 5. Correlation heatmap
```

```
sns.heatmap(df.corr(), annot=True, cmap='coolwarm')
```

```
plt.title("Feature Correlation")
```

```
plt.show()
```

```
# 6. Histogram of Fare
```

```
sns.histplot(df['fare'], bins=30, kde=True)
```

```
plt.title("Fare Distribution")
```

```
plt.xlabel("Fare")
```

```
plt.ylabel("Count")
```

```
plt.show()
```