P3_students_Questions

January 31, 2026

Practice 3: Questions

0.1 Data Loading & Exploration

Do the following: 1. Load the Titanic dataset using seaborn 2. Display the first 10 rows. 3. Count missing values in each column 4. Display column names. 5. Show unique values in the `embarked` column.

```
[1]: #Import the necessary libraries
     import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
     import numpy as np

     #To suppress all warnings
     import warnings
     warnings.filterwarnings('ignore')
```

```
[1]: # 1. Load Titanic dataset
```

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

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

```
[4]: #4.
```

```
[5]: #5.
```

0.2 Data Cleaning

Do the following: 1. Fill missing `embarked` with the most frequent value. 2. Convert `sex` and `class` to categorical type. 3. Create a new column `family_size= sibsp + parch + 1`. 4. Cap `fare` at the 99th percentile to reduce outlier impact.

```
[6]: #1.
```

```
[7]: #2.
```

```
[8]: #3.
```

```
[9]: #4.
```

0.3 Visualization

Do the following: 1. Pairplot of `['age', 'fare', 'sibsp']` colored by `survived`. 2. Heatmap of correlations for numeric columns. 3. Scatter plot of `age` vs `fare`, colored by `survived`. 4. Plot a histogram of `age` with 30 bins and KDE

```
[10]: #1.
```

```
[11]: #2.
```

```
[12]: #3.
```

```
[13]: #4.
```

```
[ ]:
```