MLOPS_Final_Project_Group15

1. Installing Required Libraries

```
!pip install neptune-client dataprep seaborn matplotlib
```

Show hidden output

2. Importing DataPrep and other functions

```
import pandas as pd
import neptune.new as neptune
import os
```

```
[neptune] [warning] NeptuneDeprecationWarning: The 'neptune-client' package has been deprecated and will be removed in the f
[neptune] [warning] NeptuneDeprecationWarning: You're importing the Neptune client library via the deprecated `neptune.new`
```

```
from dataprep.datasets import load_dataset, get_dataset_names
from dataprep.eda import create_report, plot
```

3. Listing DataPrep Datasets

```
dataset_names = get_dataset_names()
print(dataset_names)
```

```
['countries', 'iris', 'house_prices_train', 'adult', 'titanic', 'covid19', 'waste_hauler', 'patient_info', 'house_prices_tes
```

4. Loading 'titanic' Dataset and Creating EDA Report

```
df = load_dataset("titanic")
create_report(df)
```

Show hidden output

5. Saving the Created Report

```
report = create_report(df)
report.save("titanic_eda_report.html")
```

Show hidden output

```
os.listdir()
```

```
['.config', 'titanic_eda_report.html', 'sample_data']
```

6. Downloading the Created Report

```
from google.colab import files
files.download("titanic_eda_report.html")
```

7. Plotting 'Age' Variable to study plot() function

```
plot(df, 'Age')
```

Stats | Histogram | KDE Plot | Normal Q-Q Plot | Box Plot | Value Table

Overview

Approximate Distinct Count	88
Approximate Unique (%)	12.3%
Missing	177
Missing (%)	19.9%
Infinite	0
Infinite (%)	0.0%
Memory Size	11424
Mean	29.6991
Minimum	0.42
Maximum	80
Zeros	0
Zeros (%)	0.0%
Negatives	0
Negatives (%)	0.0%

Descriptive Statistics

Mean	29.6991
Standard Deviation	14.5265
Variance	211.0191
Sum	21205.17
Skewness	0.3883
Kurtosis	0.1686
Coefficient of Variation	0.4891

Quantile Statistics

Minimum	0.42
5-th Percentile	4
Q1	20.125
Median	28
Q3	38
95-th Percentile	56
Maximum	80
Range	79.58
IQR	17.875

8. Initializing Neptune.ai Run

```
run = neptune.init_run(
    project='sbonga4/MLOPS-Final-Project-Group15',
    api_token='eyJhcGlfYWRkcmVzcyI6Imh0dHBzOi8vYXBwLm5lcHR1bmUuYWkiLCJhcGlfdXJsIjoiaHR0cHM6Ly9hcHAubmVwdHVuZS5haSIsImFwaV9rZXkiC
)
```

Show hidden output

9. Importing Matplotlib and Seaborn for Custom Visualizations

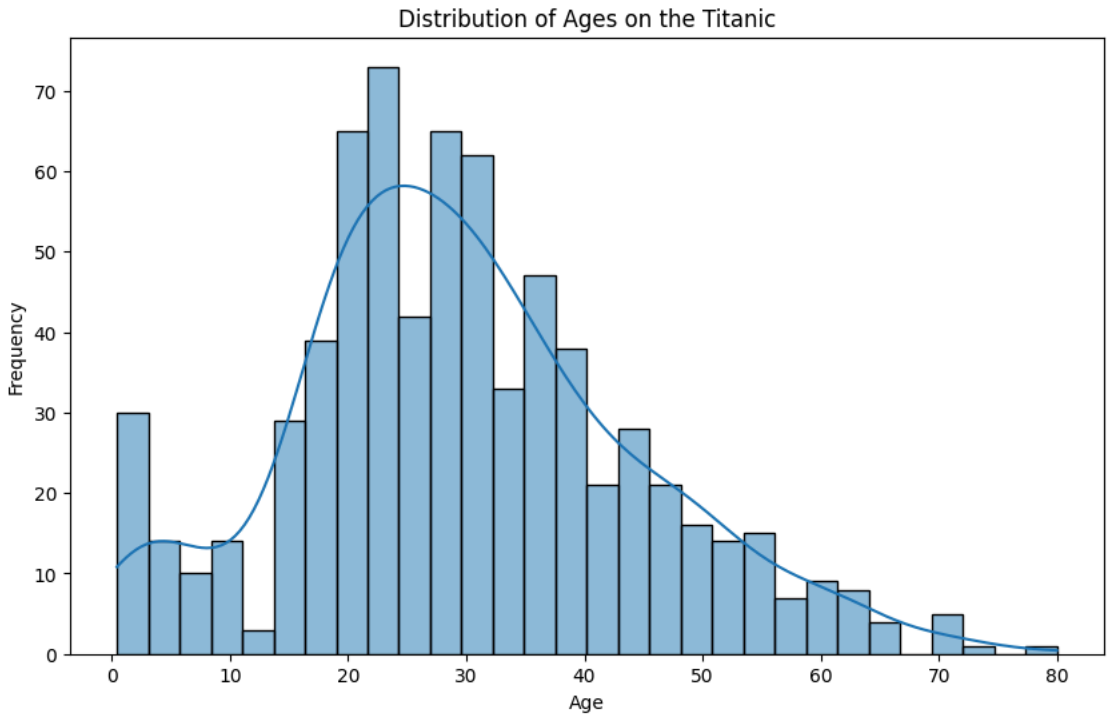
```
import seaborn as sea
import matplotlib.pyplot as mat
```

10. Plots using Matplotlib and Seaborn

```
mat.figure(figsize=(10, 6))
sea.histplot(df['Age'].dropna(), bins=30, kde=True)
mat.title('Distribution of Ages on the Titanic')
mat.xlabel('Age')
mat.ylabel('Frequency')

mat.savefig('titanic_age_distribution.png')

mat.show()
```




```
#df_filtered = df[(df['Age'] >= 25) & (df['Age'] <= 45)]
#df_filtered = df[(df['Age'] >= 15) & (df['Age'] <= 40) & (df['Sex'] == 'female')]
df_filtered = df[(df['Age'] >= 20) & (df['Age'] <= 30) & (df['Sex'] == 'male')]


#mat.figure(figsize=(10, 6))
#sea.histplot(df_filtered['Age'].dropna(), bins=30, kde=True)
#mat.title('Distribution of Ages (25 to 45) on the Titanic')
#mat.xlabel('Age)')
#mat.ylabel('Frequency')
#mat.savefig('titanic_age_distribution_25_to_45.png')
#mat.show()

#mat.figure(figsize=(10, 6))
#sea.histplot(df_filtered['Age'].dropna(), bins=30, kde=True)
#mat.title('Distribution of Females Aged (15 to 40) on the Titanic')
#mat.xlabel('Age)')
#mat.ylabel('Frequency')
#mat.savefig('titanic_female_age_distribution_15_to_40.png')
#mat.show()

mat.figure(figsize=(10, 6))
sea.histplot(df_filtered['Age'].dropna(), bins=30, kde=True)
mat.title('Distribution of Males Aged (20 to 30) on the Titanic')
mat.xlabel('Age)')
mat.ylabel('Frequency')
mat.savefig('titanic_male_age_distribution_20_to_30.png')
mat.show()
```

Show hidden output

11. Uploading the Visualizations to Neptune and Stopping the Run

```
run['visualizations/titanic_age_distribution_chart'].upload('titanic_male_age_distribution_20_to_30.png')
run.stop()
```

```
[neptune] [info   ] Shutting down background jobs, please wait a moment...
[neptune] [info   ] Done!
[neptune] [info   ] Waiting for the remaining 1 operations to synchronize with Neptune. Do not kill this process.
[neptune] [info   ] All 1 operations synced, thanks for waiting!
[neptune] [info   ] Explore the metadata in the Neptune app: https://app.neptune.ai/sbonga4/MLOPS-Final-Project-Group15/e/ML
```

Start coding or generate with AI.