DATAPREP AI AND MISSING NO

WITH PYTHON

DR. ALVIN ANG

CONTENTS

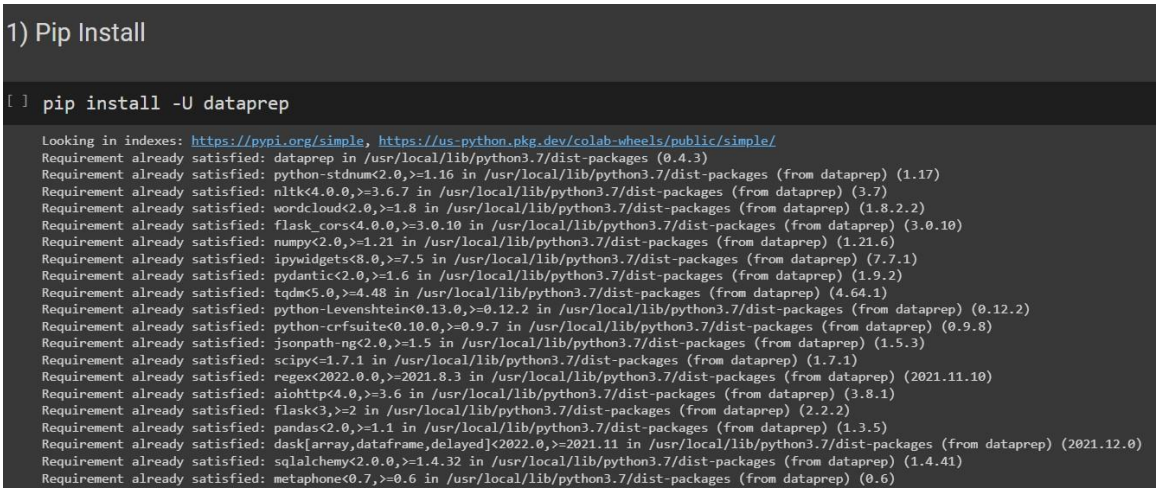
I. PIP INSTALL

https://www.alvinang.sg/s/DataPrep__MissingNo_by_Dr_Alvin_Ang.ipynb

References:

https://towardsdatascience.com/using-the-missingno-python-library-to-identify-and-visualise-missing-data-prior-to-machine-learning-34c8c5b5f009#:~:text=Missingno%20is%20an%20excellent%20and,%2C%20heatmap%2C%20or%20a%20dendrogram

https://dataprep.ai/



II. IMPORT LIBRARIES

A. IMPORT DATAPREP

2) Import Libraries

2a) Import DataPrep

```
[ ] from dataprep.datasets import load_dataset
    from dataprep.eda import plot, plot_correlation, plot_missing, plot_diff, create_report
```

B. IMPORT MISSINGNO

2b) Import MissingNo

```
[ ] import pandas as pd
    import missingno as msno
    import matplotlib.pylab as plt
    %matplotlib inline
```

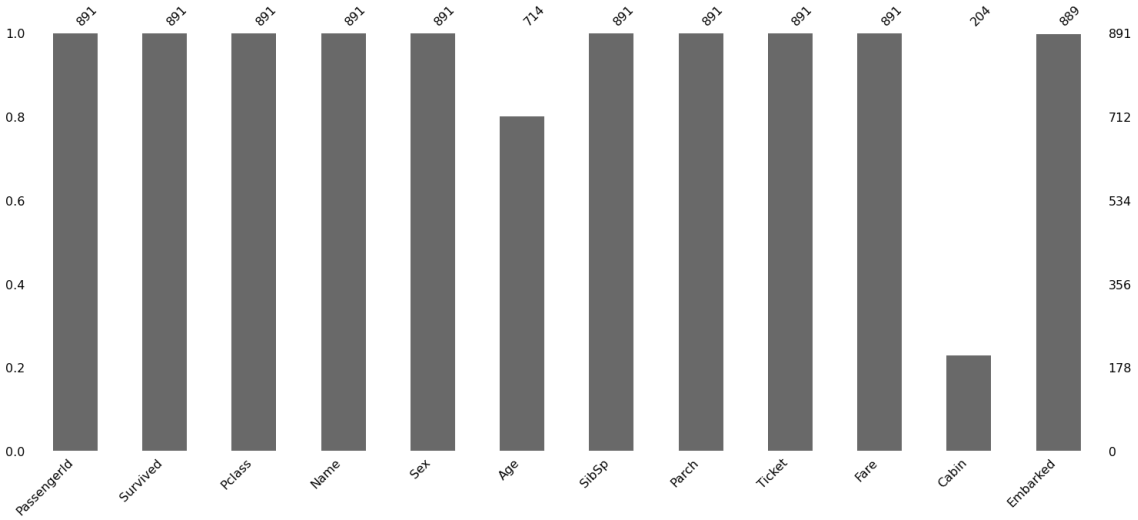
C. LOAD DATASET

2c) Load Dataset

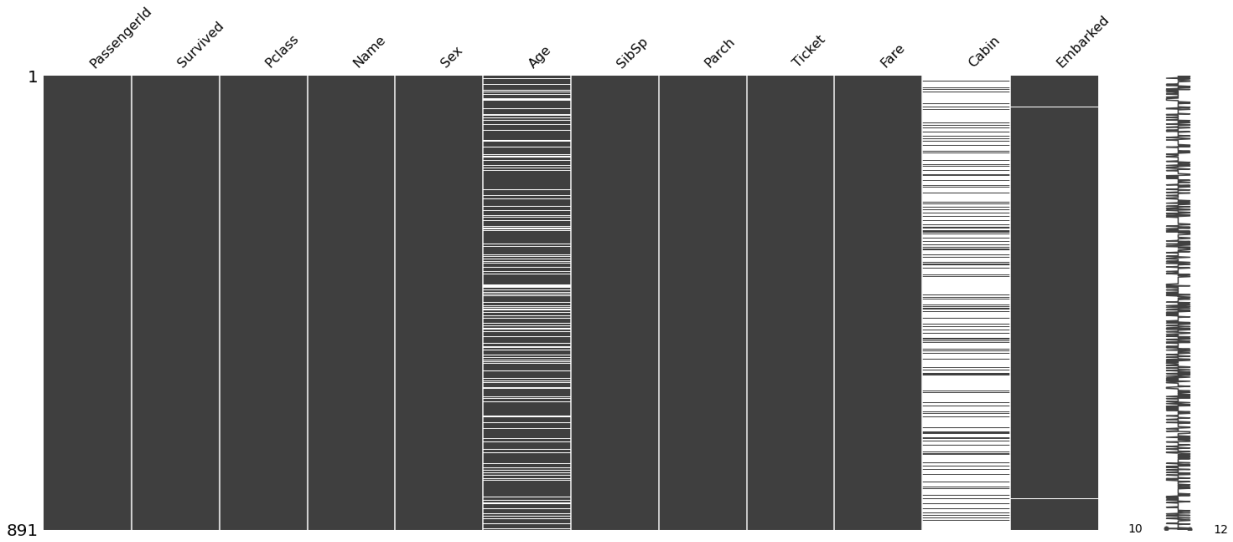
```
[ ] df = load_dataset("titanic")
```

3) MSNO

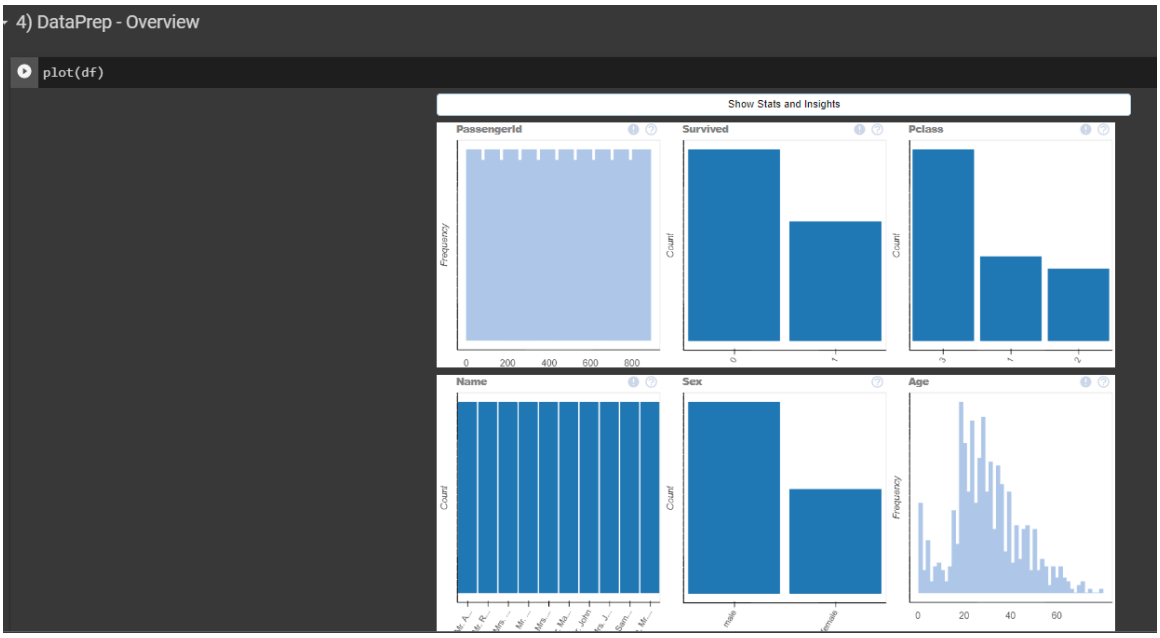
```
msno.bar(df)
```

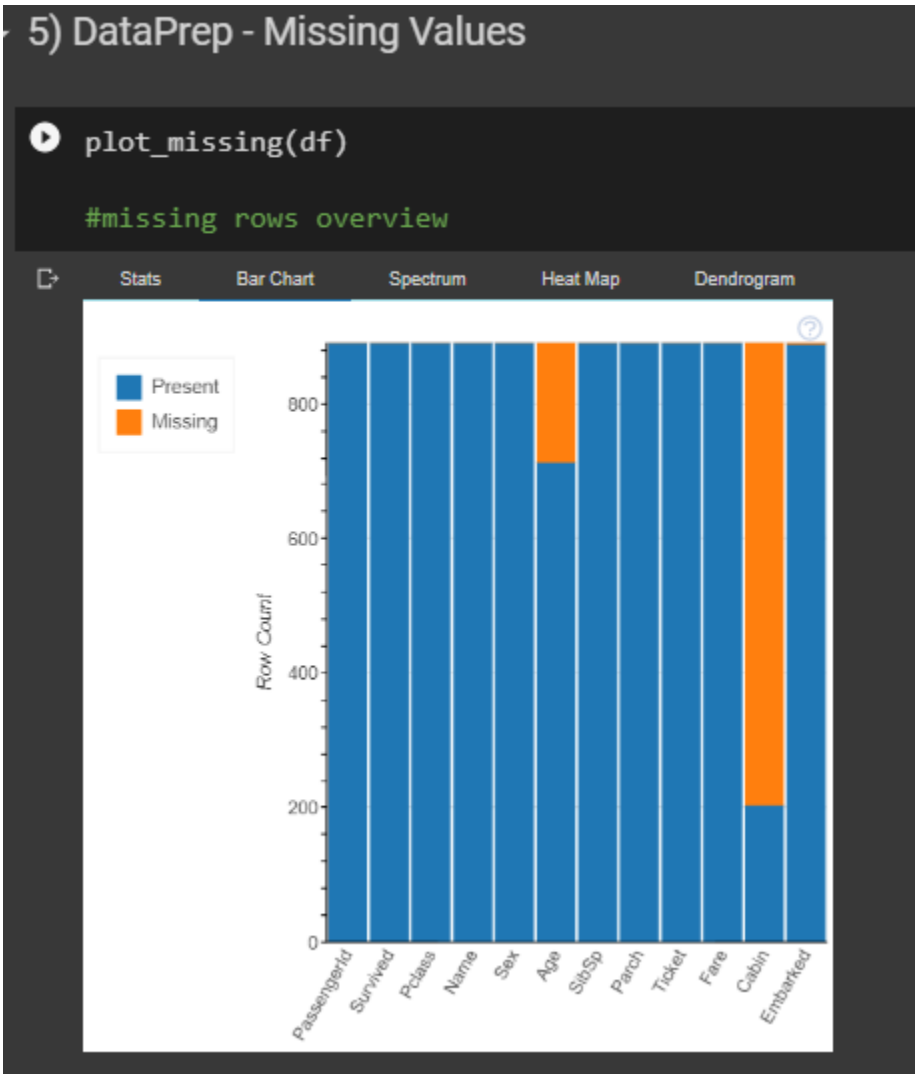


```
msno.matrix(df)
```

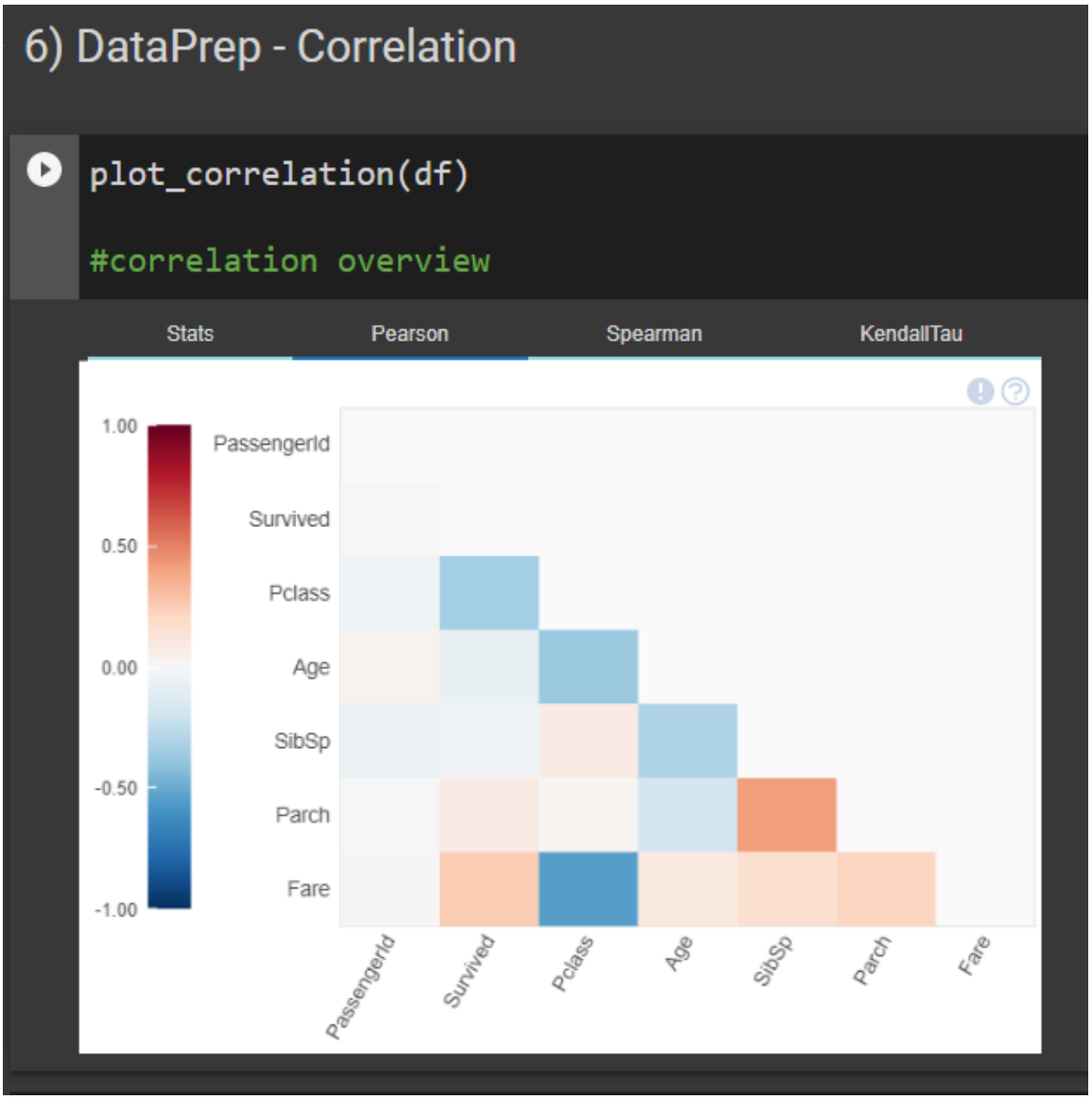


IV. DATAPREP - OVERVIEW

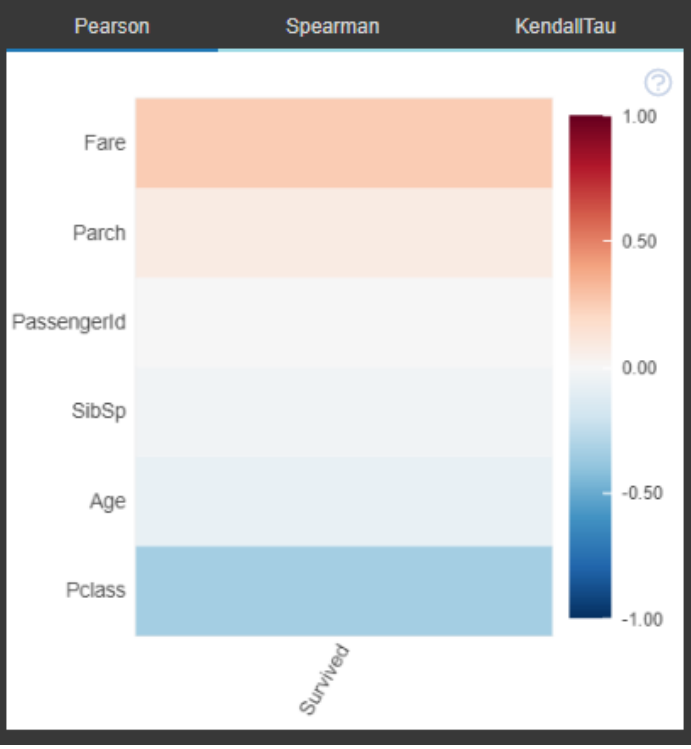


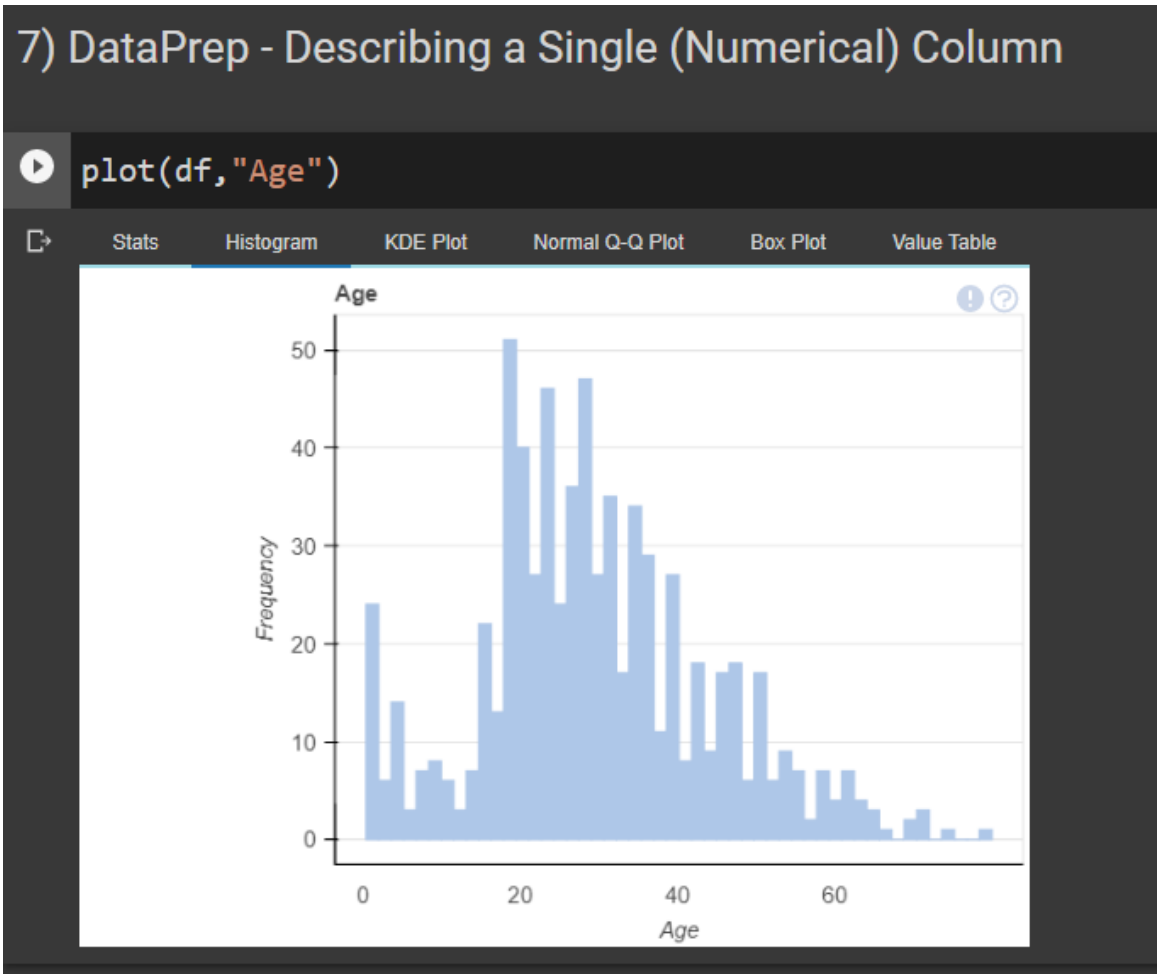


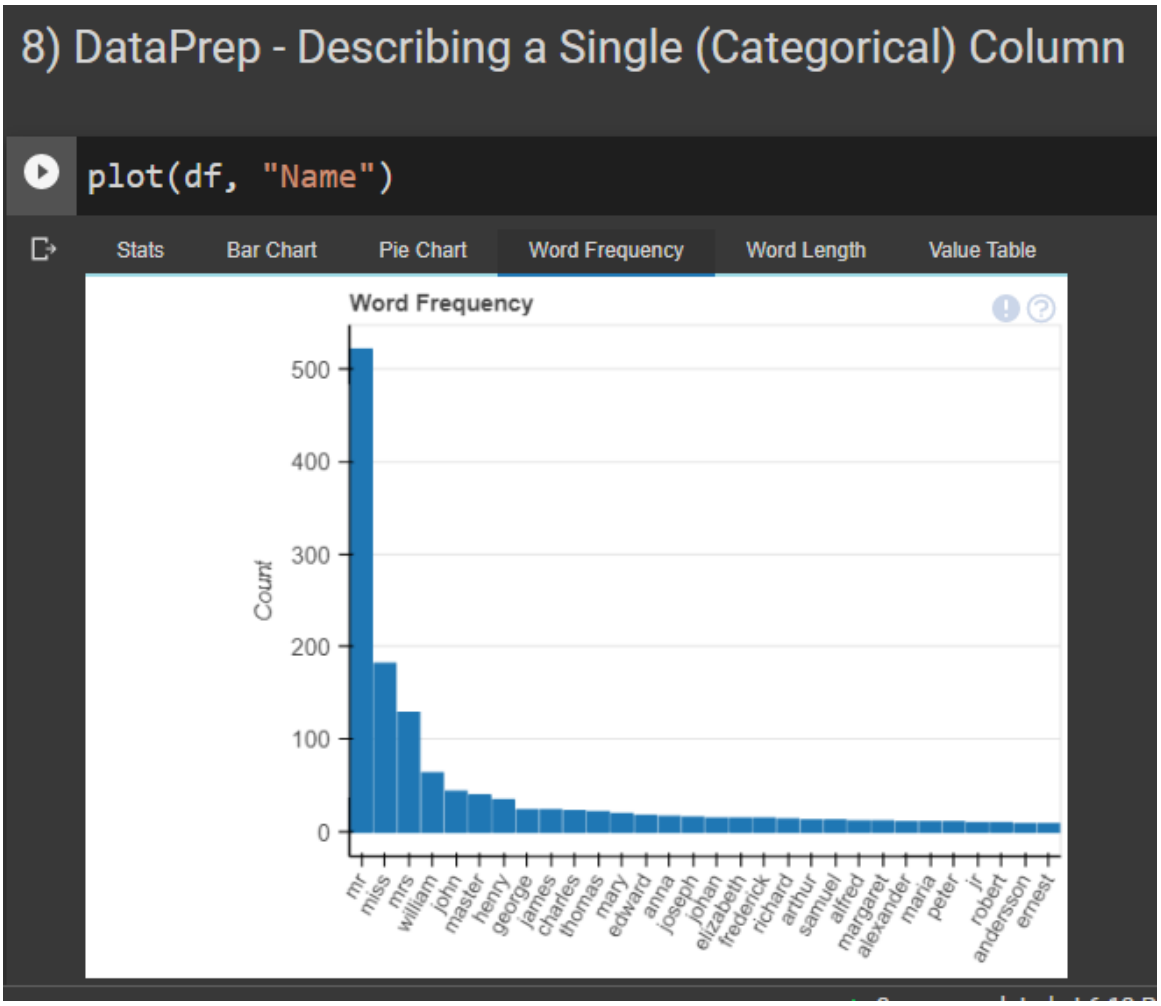
```
[ ] plot_missing(df, 'Age ')

    #understand how the other columns change
    #after dropping the missing values of 'age' column
```

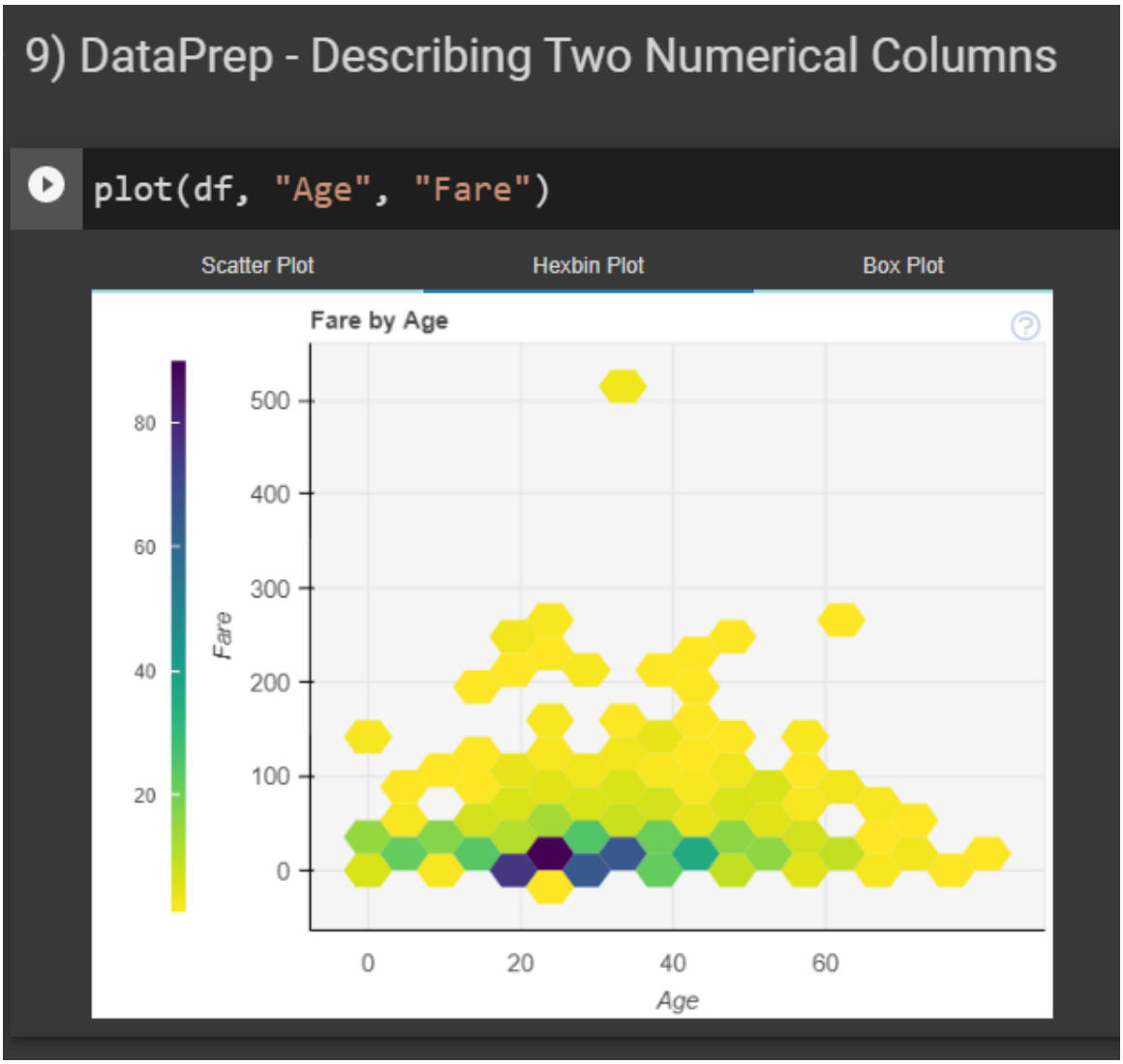


```
plot_correlation(df, "Survived")

#understand how other columns correlate to "Survived" column
```

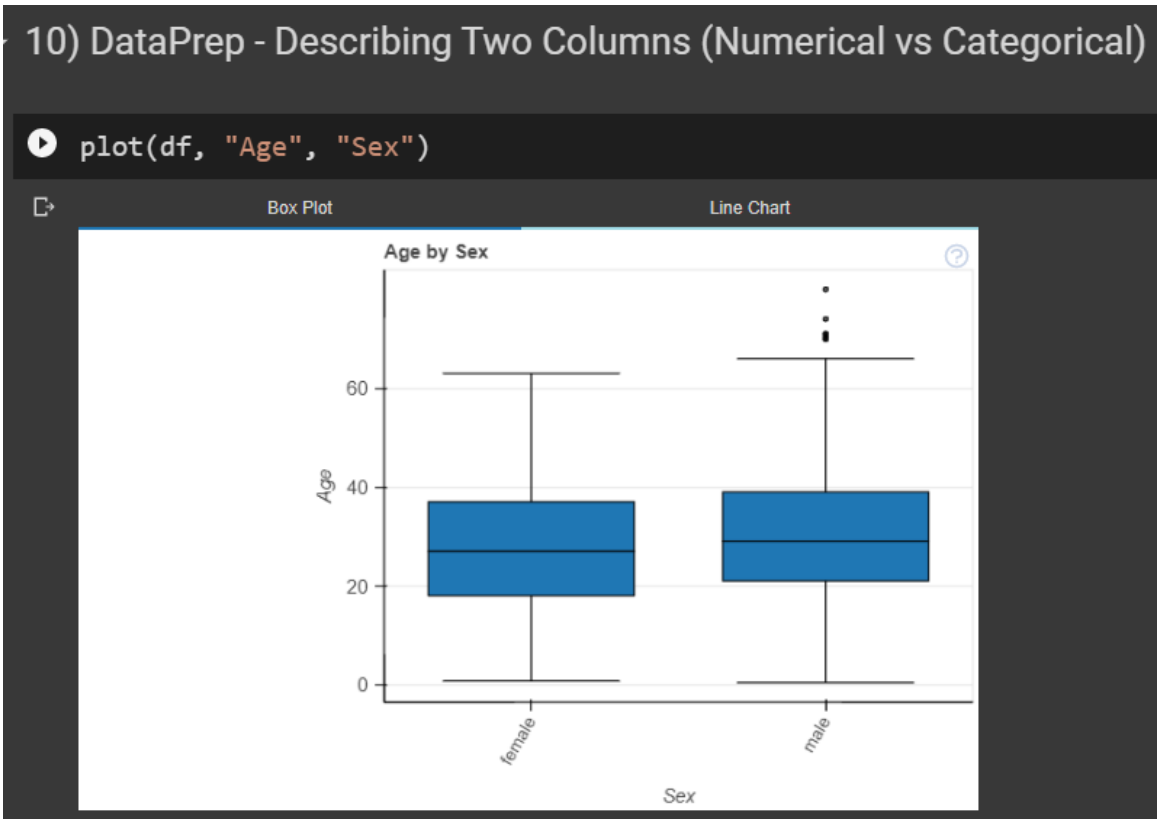


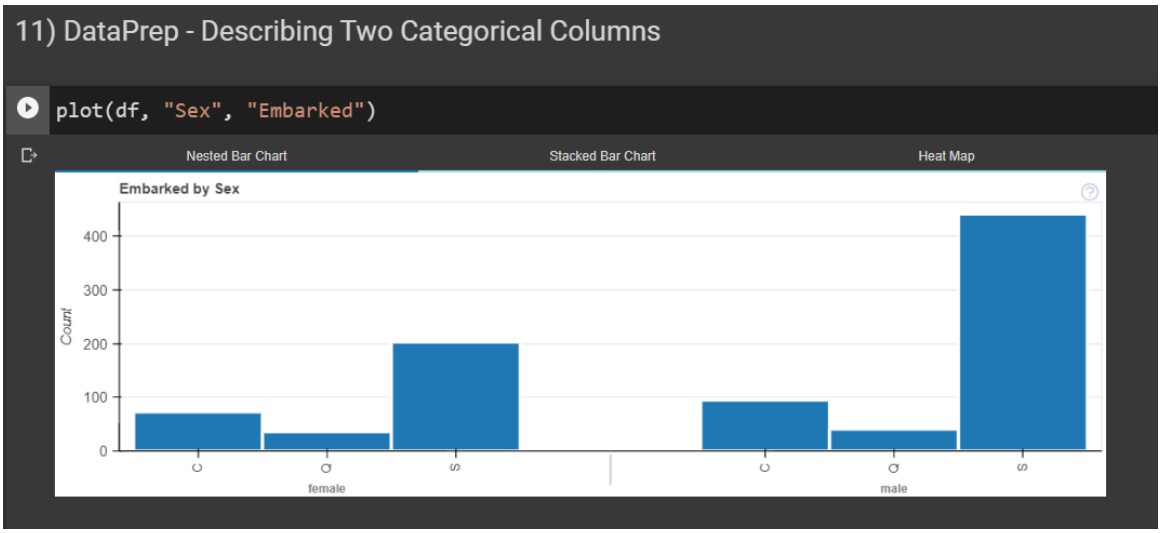






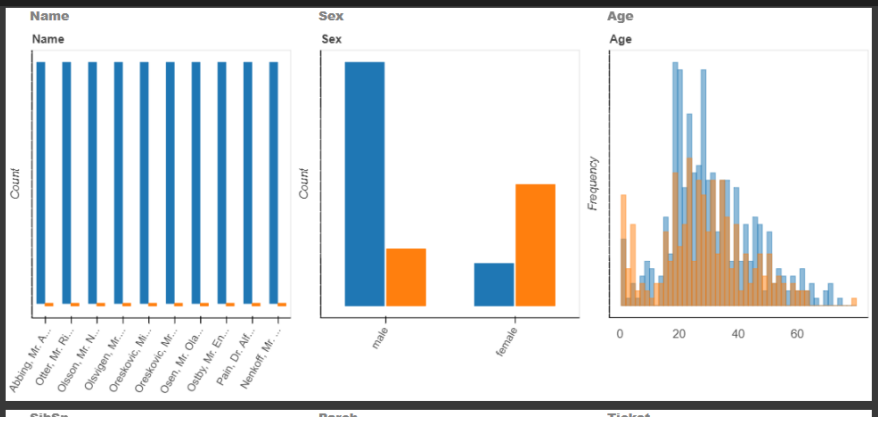
X. DATAPREP – DESCRIBING TWO COLUMNS (NUMERICAL VS CATEGORICAL)

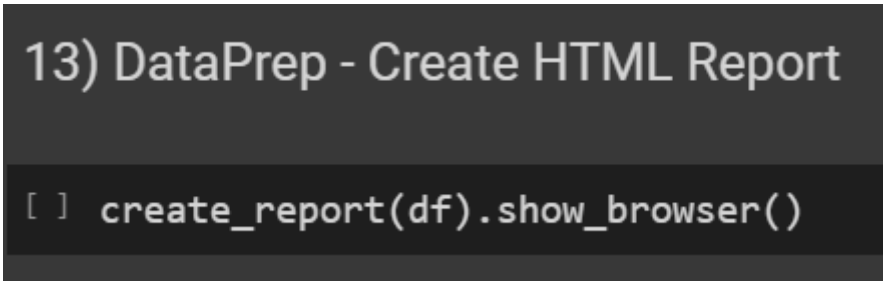




12) DataPrep - Compare Two Dataframes

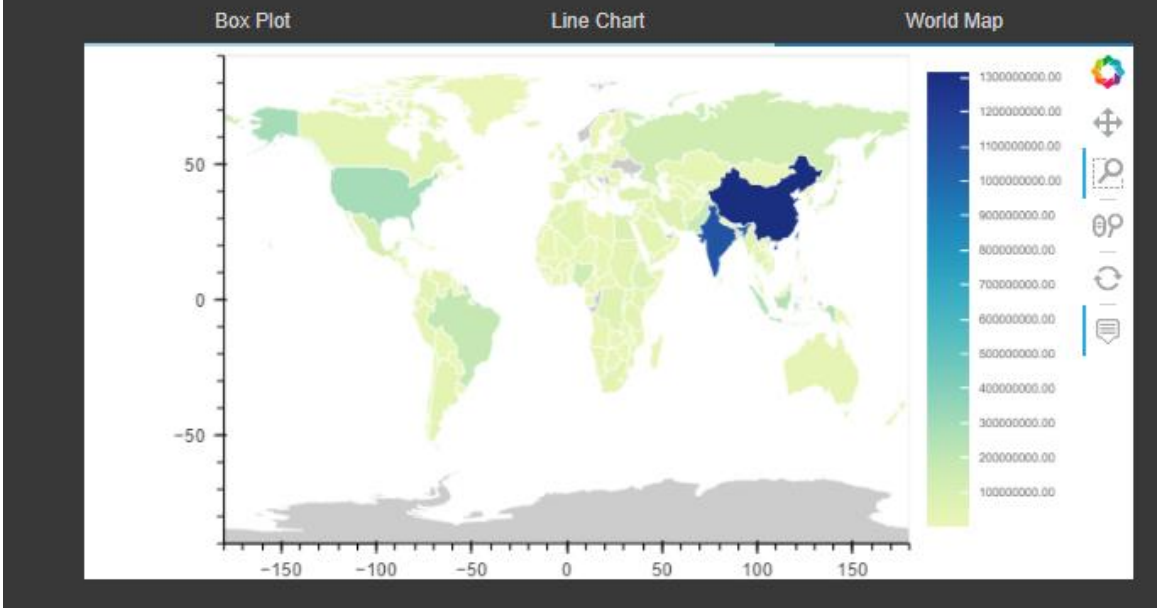
```
df1 = df[df["Survived"] == 0]
df2 = df[df["Survived"] == 1]
plot_diff([df1, df2])
```





14) DataPrep - Analyze Geographical Data

```
[ ] country = load_dataset("countries")
    plot(country, "Country", "Population")
```



ABOUT DR. ALVIN ANG

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He is a scientist, entrepreneur, as well as a personal/business advisor. More about him at www.AlvinAng.sg.