Assignment

Use the "from the expert" (FTE) jupyter notebook as a starter for this assignment, and ask your instructor questions if you need help.

Use the `churn_data.csv` file to carry out a similar EDA and visualization process as what we did in the FTE. Create at least 2 EDA plots, and create a HTML file with an auto-EDA analysis using pandas-profiling or another auto-EDA Python package. Write a short analysis at the end of the assignment in markdown.

Data science process steps this week

We will carry out the first two parts of the CRISP-DM data science process this week:

1. Business understanding

This is customer churn data for a telecommunications company. Customers can have phone as well as other services. The company is looking to reduce customer churn, where customers stop using the company's services and cancel their account. The 'Churn' column has a binary target, yes or no, that denotes if a customer churned. We want to create a machine learning model to predict the Churn target using the other available data in the dataset. Ideally, we will deploy this model to integrate with the company's database, so that a churn risk column is created for each customer. This will enable customer service reps and others to devise and use strategies to reduce churn.

2. Data understanding

Carry out some EDA as we did in the FTE, such as using pandas-profiling. Create a histogram like we did in the FTE, where we plot a numeric column with the target as the 'hue'. Optional challenge: create other plots with the target as the hue, such as bar plots for the categorical columns.

[2]:
```
import pandas as pd
from pandas_profiling import ProfileReport
# %matplotlib inline
import matplotlib.pyplot as plt
%matplotlib inline
```

[3]:
```
df = pd.read_csv('churn_data.csv')#, index_col='Churn')
df
```

[3]:

	customerID	tenure	PhoneService	Contract	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	1	No	Month-to-month	Electronic check	29.85	29.85	No
1	5575-GNVDE	34	Yes	One year	Mailed check	56.95	1889.50	No
2	3668-QPYBK	2	Yes	Month-to-month	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	45	No	One year	Bank transfer (automatic)	42.30	1840.75	No
4	9237-HQITU	2	Yes	Month-to-month	Electronic check	70.70	151.65	Yes
...	...	...	...	...	...	...	...	...
7038	6840-RESVB	24	Yes	One year	Mailed check	84.80	1990.50	No
7039	2234-XADUH	72	Yes	One year	Credit card (automatic)	103.20	7362.90	No
7040	4801-JZAZL	11	No	Month-to-month	Electronic check	29.60	346.45	No
7041	8361-LTMKD	4	Yes	Month-to-month	Mailed check	74.40	306.60	Yes
7042	3186-AJIEK	66	Yes	Two year	Bank transfer (automatic)	105.65	6844.50	No

7043 rows × 8 columns

[4]:
```
report = ProfileReport(df)
report.to_file('churn_eda.html')
```

Summarize dataset: 100% 31/31 [00:08<00:00, 4.10it/s, Completed]

Generate report structure: 100% 1/1 [00:03<00:00, 3.61s/it]

Render HTML: 100% 1/1 [00:01<00:00, 1.20s/it]

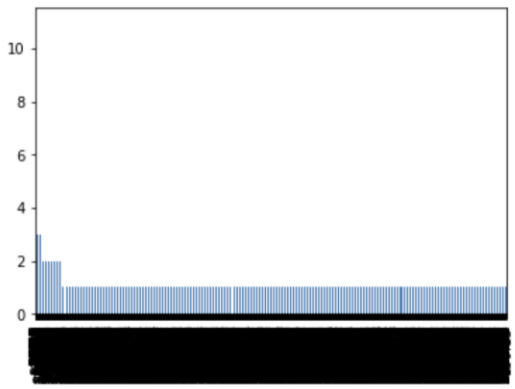
Export report to file: 100% 1/1 [00:00<00:00, 16.06it/s]

[5]:
```
df['TotalCharges']
#type(df['TotalCharges'])
```

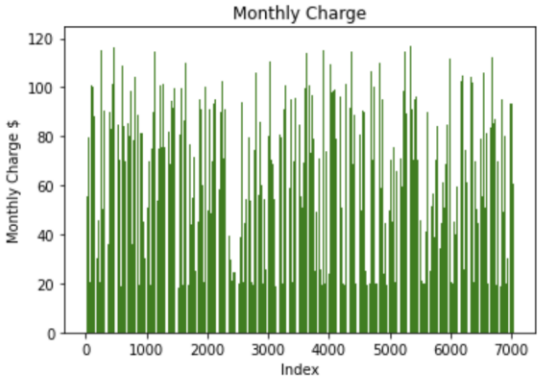
[5]:
```
0         29.85
1       1889.50
2        108.15
3       1840.75
4        151.65
          ...
7038    1990.50
7039    7362.90
7040     346.45
7041     306.60
7042    6844.50
Name: TotalCharges, Length: 7043, dtype: float64
```

[6]:
```
df['TotalCharges'].value_counts().plot.bar()
```

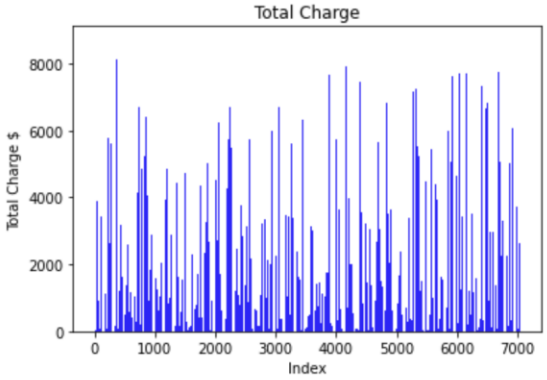
[6]:
```
<AxesSubplot:>
```



[7]:
```
dframe = pd.DataFrame(df)
monChg = list(dframe.iloc[:,5])
totChg = list(dframe.iloc[:,6])
index = range(len(monChg))
#monChg
#totChg
plt.bar(index, monChg, color='g')
plt.title("Monthly Charge")
plt.xlabel("Index")
plt.ylabel("Monthly Charge $")
plt.show()
```



[8]:
```
plt.bar(index, totChg, color='b')
plt.title("Total Charge")
plt.xlabel("Index")
plt.ylabel("Total Charge $")
plt.show()
```



[9]:
```
import markdown
output = markdown.markdown('''
#### Due 20220117, January 17, 2022
# EDA Plot 1
* I created an EDA plot that plotted a bar chart of the Monthly Charge
* The Monthly Charge is the money paid by a customer each month.
# EDA Plot 2
* I created an EDA plot that plotted a bar chart of the Total Charge
* The Total Charge is the total money paid by a customer to the company.
# Auto-EDA Analysis HTML
* I used pandas-profiling to create an analysis exported to an HTML file.
#### Thank you!
#### Created by Jeremy Beard
#### Regis University
#### MSDS 600
''')

print(output)
file = open('markdown.html', 'w')
print(output, file = file)
```