Python Data Analysis Project Instructions

1. Download Anaconda if you have not already so you can use Jupyter Notebook
2. Find a dataset of your interest from Tableau Public, Kaggle, or Google Dataset Search and import it into python using one of the pandas functions (read_csv, read_excel)
3. Think of 2-3 business questions that you'd like to answer using the data
   (1. What does the titanic data set look like and how are they categorized?
   2. What does the data look like for survivors?
   3. Was there any correlations made between comparing the data?)
4. Write the code and document your steps in a Jupyter notebook to accomplish the following:

   Load your data:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sb
```

   a. **Clean your data for analysis and get it into DataFrames**

```
dataset = pd.read_csv('/Users/michael/Desktop/titanic.csv', index_col = 2)
```

   b. **Demonstrate the following in your notebook:**

      i. **Ability to index DataFrames**

```
dataset.set_index('Sex')
```

      ii. **Ability to view column headers, first n rows, and last n rows in a DataFrame**

```
dataset.head(10)
dataset.tail(10)
```

      iii. **Ability to select specific data points from a DataFrame**

```
dataset.drop(['Pclass','Siblings/Spouses Aboard','Parents/Children Aboard','Survived'],axis=1).head(21)
```

      iv. **Ability to sort DataFrames**

```
dataset.sort_values(by='Age')
```

v. **Ability to summarize DataFrames using GroupBy**
dataset.groupby(['Fare','Survived']).mean()

vi. **Ability to run basic descriptive statistics on a DataFrame**
dataset.describe()

vii. **Ability to build a basic chart with DataFrames with plot()**
pd.value_counts(dataset['Sex'] ,['Survived']).plot.bar()

c. Run a basic data analysis to find at least two interesting insights and describe why they are useful
d. You should call out the code that produces these insights, as well as a visual representation of the insights (visualization, table, number, etc.)
e. You should use comments/markup to describe the insight and why it's important directly in the Jupyter Notebook

Submission

1. Post your code on Github
2. Present your analysis and findings directly from your Jupyter Notebook
3. Create 3-5 minutes of a recorded video presentation (using Loom) that explains your final project (business questions, dataset, overview of your code, insights, and visuals). Your video should not go over 5 minutes. Create a succinct video that can effectively communicate your message in a short time to busy business users.