

PDS Project FAQ's

1. How to submit the answers to the questions asked in the project?

Ans: Write answer in the Jupyter notebook, & show the Output their itself. Also, explain the answers/inferences as per the question either in markdown or print it.

2. Library not found?

Ans: There might be a situation where you get an error message similar to the one mentioned below:

```
In [1]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
sns.set_color_codes=True
%matplotlib line

In [3]: df=pd.read_csv("C:/Users/admin/Downloads/RAJ/ForbesAmericasTopColleges2019.csv")
df.head()

-----
NameError                                Traceback (most recent call last)
<ipython-input-3-de0dd9db1be7> in <module>
----> 1 df=pd.read_csv("C:/Users/admin/Downloads/RAJ/ForbesAmericasTopColleges2019.csv")
      2 df.head()

NameError: name 'pd' is not defined
```

In this case, try to uninstall & install again that particular library. Use the below set of codes to run:

pip uninstall pandas -- yes

pip install pandas

Import pandas as pd

pip uninstall numpy -- yes

pip install numpy

Import numpy as np

pip uninstall matplotlib -- yes

pip install matplotlib

Import matplotlib as plt

```
pip uninstall seaborn -- yes
```

```
pip install seaborn
```

```
Import seaborn as sns
```

Note: Kindly put double hyphen(-) before yes.

3. Warning message?

Ans: There might be a situation where you receive the below warning message; as of now you can ignore it. This message basically makes us aware that there may be some function which we are using currently will be removed from updated/new versions. Also, in the warning message itself, the new functions have been defined which can be used.

```
C:\Users\Admin\anaconda3\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `kdeplot` (an axes-level function for kernel density plots).
warnings.warn(msg, FutureWarning)
```

4. Should we drop NA values completely from the original dataset? (Question-9)

Ans: You should first copy the data in "df" object and then drop NA values from the "df".

5. In START* and STOP* columns, there is a value as R?walpindi and Rawalpindi. While analysing the unique values, these are treated as different values. Is it ok to proceed treating these as two different values? (Question-11)

Ans: For this project, please treat R?walpindi and Rawalpindi as different entities. You will learn how to clean this type of issue in the future courses

6. "Unknown Location" can be many, so should we consider this in Unique Start Location or Unique Stop Location or we should remove this from Unique Location? (Question -11)

Ans: Hint, don't consider NA values

7. Do we need to display the entire list or only one entry which has maximum value? (Question 15 &16)

Ans: You can choose to display the entire output in the descending order (Entry with the maximum value at the top) or only the most popular point

8. Do we need to display the entire list of most frequent routes? (Question-17)

Ans: You may choose to display only the most frequent route

9. List the most frequent route taken by Uber drivers. What is the meaning of the most frequent route and an approach to solve this question? (Question-17)

Ans: Most Frequent route means the route with most frequent START and STOP combination.

Example

Start Stop

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited.

Hyderabad Delhi

Delhi Chennai

Chennai Bengaluru

Hyderabad Bengaluru

Delhi Chennai

Delhi Bengaluru

The expected answer:

Frequent Route is Delhi -> Chennai

- 10.** Plot a bar graph of Purposes vs Miles(Distance). In this question can I take the X-axis as Purposes as the title names are overlapping? **(Question-19)**

Ans: Yes, you are free to choose either Purpose or Miles on X-axis

- 11.** Display a dataframe of Purposes and the distance travelled for that particular Purpose. What is expected from this question? **(Question-20)**

Ans: Display the total distance (using Miles variable) for each given purpose

- 12.** What is the difference between a proportion and percentage? **(Question-22)**

Ans: Proportion is basically a ratio and to if we multiply proportion with 100, we will have value in percentage

