Predictive Analytics

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Individual Coursework

Examples and indicative marks

PLAN

UCL
SCHOOL OF
MANAGEMENT

1. Visualizations
2. Narrative
3. Methodology
4. Code

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

DATASET

UCL
SCHOOL OF
MANAGEMENT

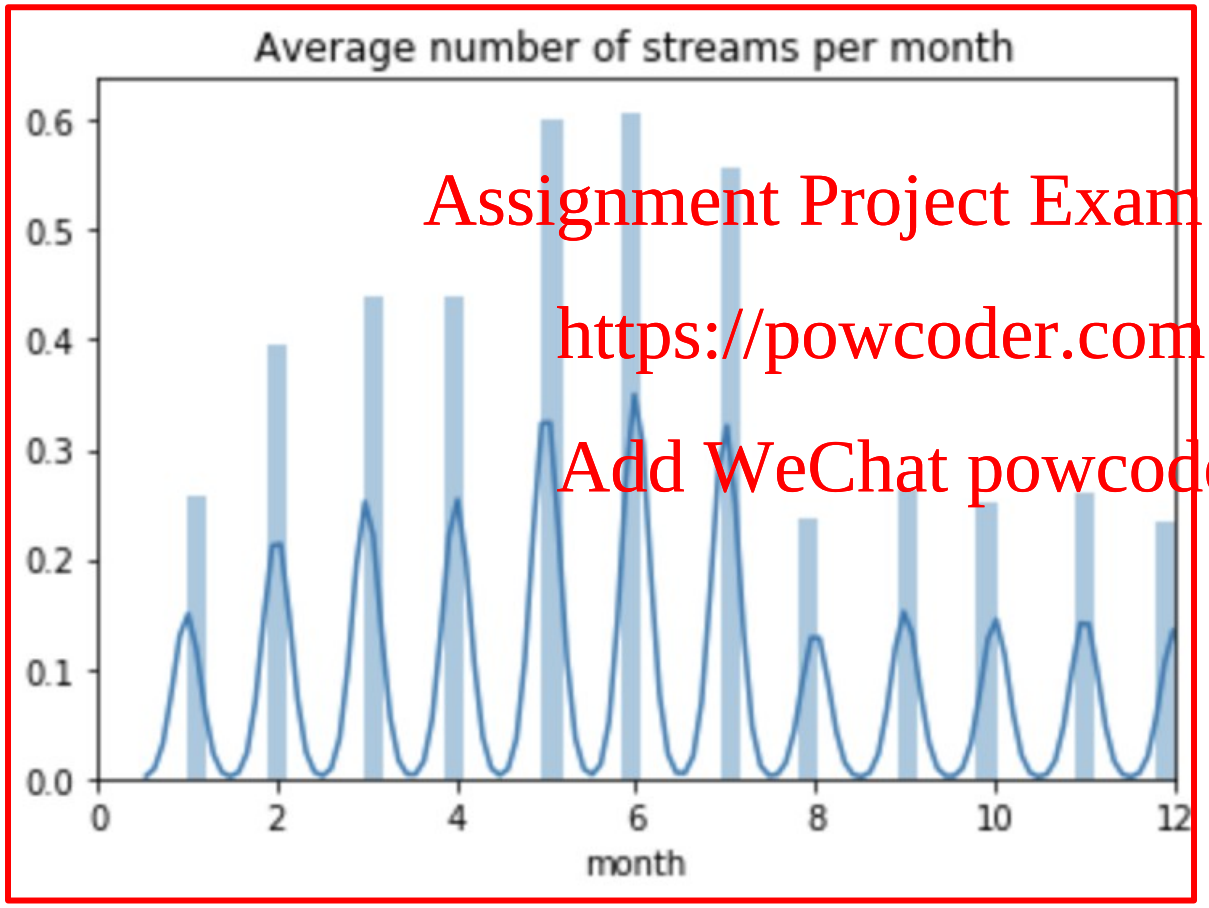
Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

APPROPRIATE VISUALIZATION

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

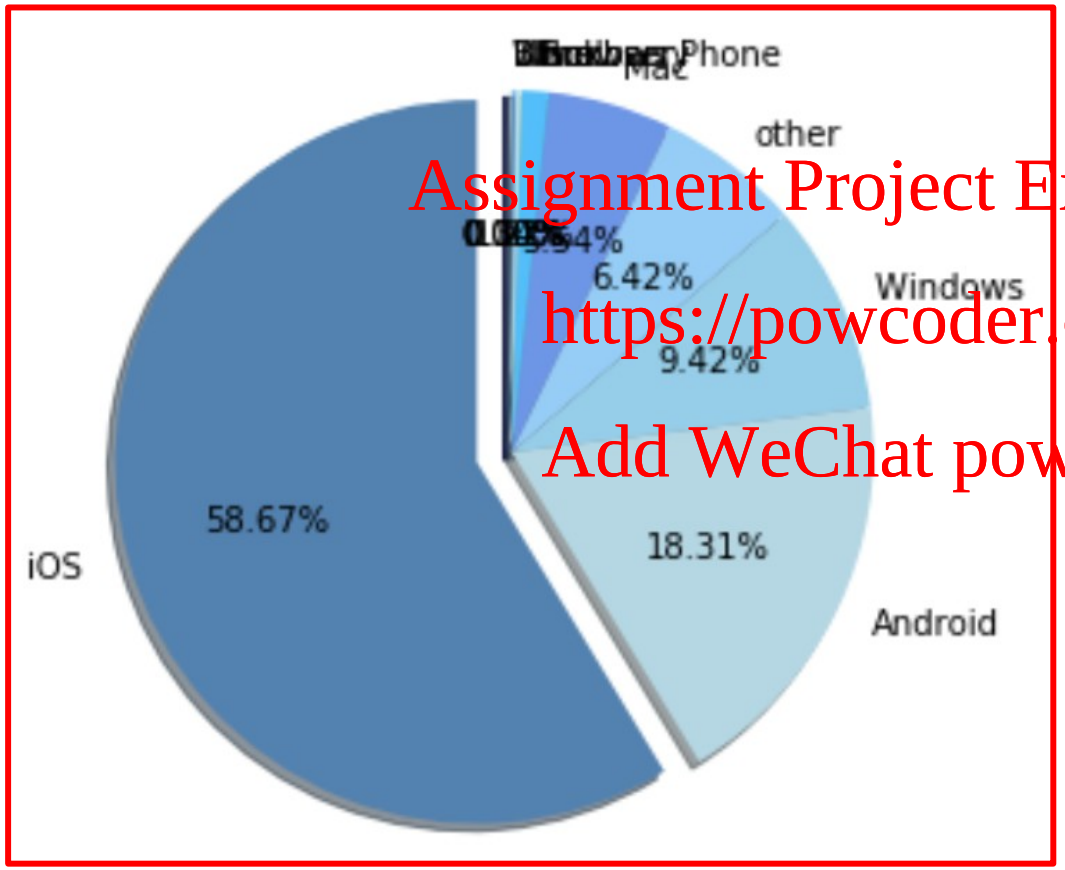
Comment:
Poor visualization for a discrete variable. Confusing.

Indicative mark: 4/10

Rubric:
Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

APPROPRIATE VISUALIZATION

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

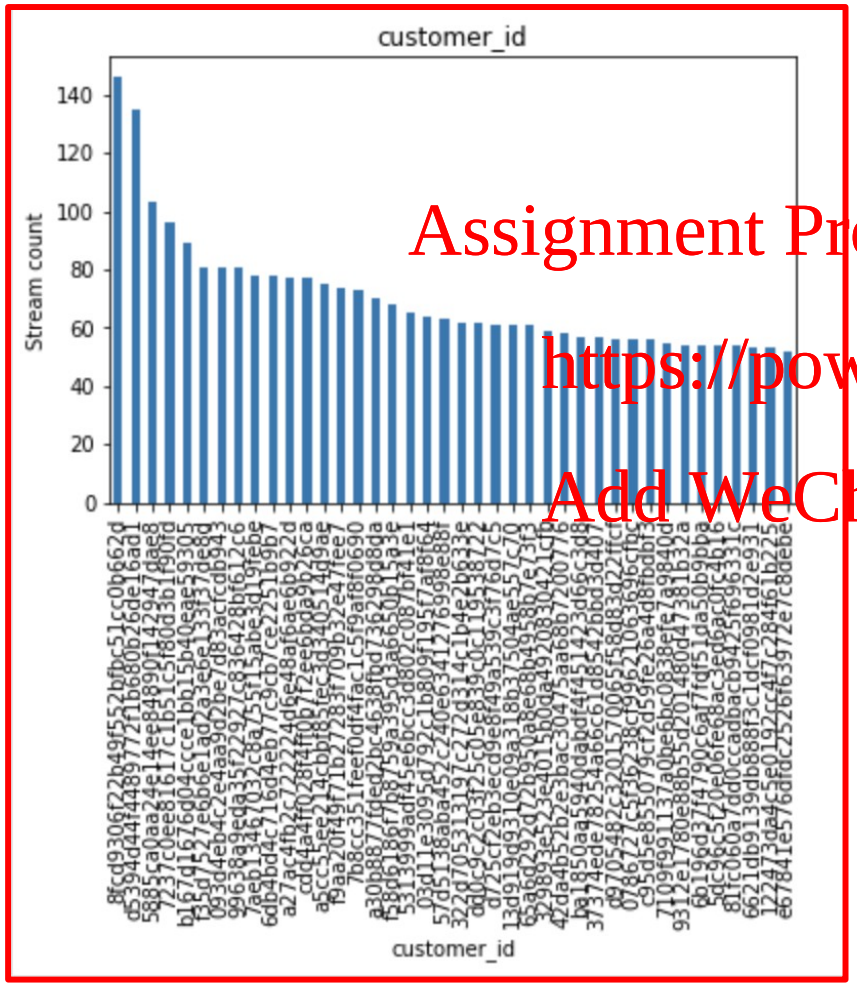
Comment:
Poor visualization for a discrete variable. Confusing, percentages and legend is not interpretable. No title.

Indicative mark: 4/10

Rubric:
Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

APPROPRIATE VISUALIZATION

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

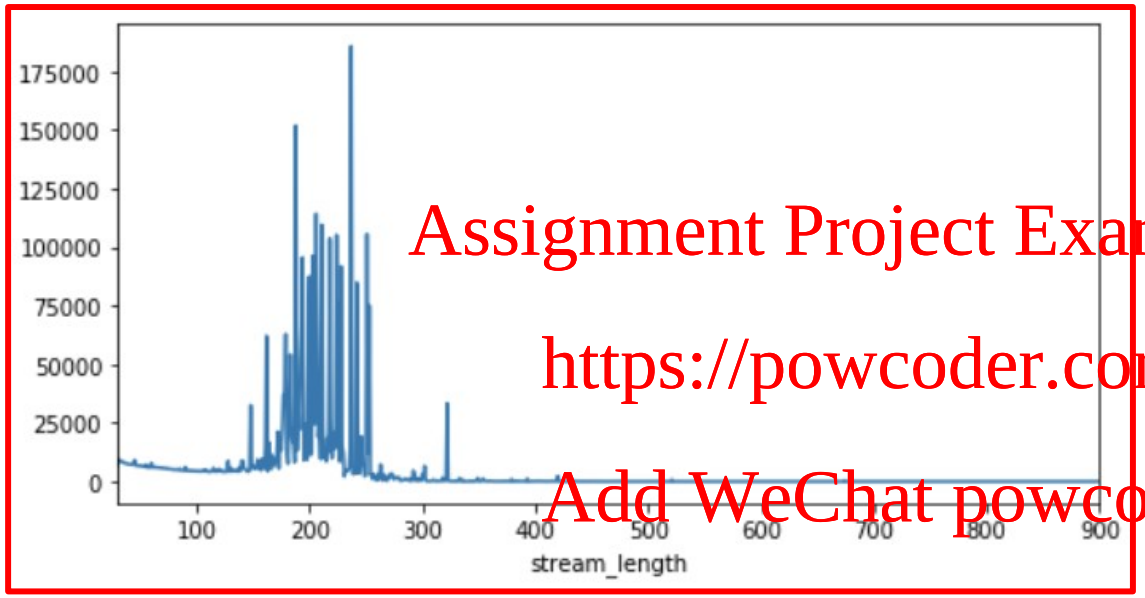
Comment: Poor visualization given the labels on the x-axis are so long; hard to see the point because of the hectic labels.

Indicative mark: 4/10

Rubric: Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

APPROPRIATE VISUALIZATION

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

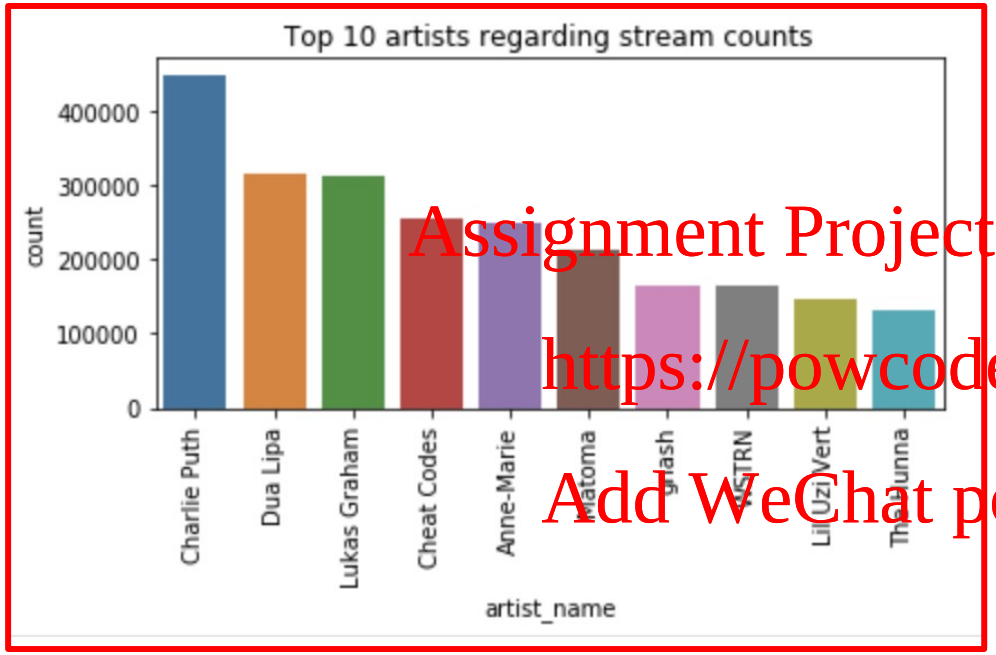
Comment: Poor visualization for this type of variable – bar chart would make more sense. Lacks title and description of the y axis.

Indicative mark: 5/10

Rubric: The report generally has a clear structure, although it would benefit from further development. Some errors and inconsistencies.

DETAILS

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

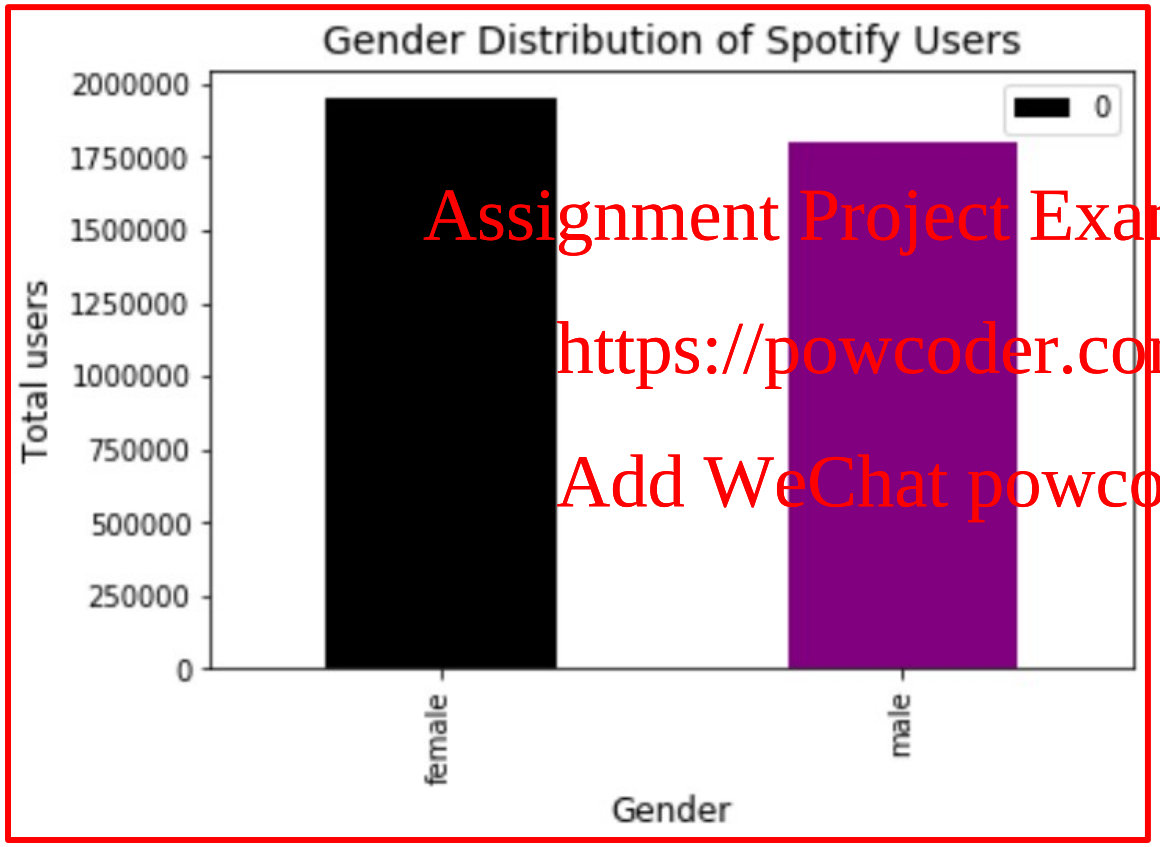
Comment: Using multiple colors makes this chart harder to interpret. Try to use different colors when it helps the reader extract more information.

Indicative mark: 6/10

Rubric: Generally well-structured and solid in appearance and usability. Some minor errors and inconsistencies in places.

DETAILS

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: The legend should have been removed.

Indicative mark: 6/10

Rubric: Generally well-structured and solid in appearance and usability. Some minor errors and inconsistencies in places.

DETAILS

UCL
SCHOOL OF
MANAGEMENT

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

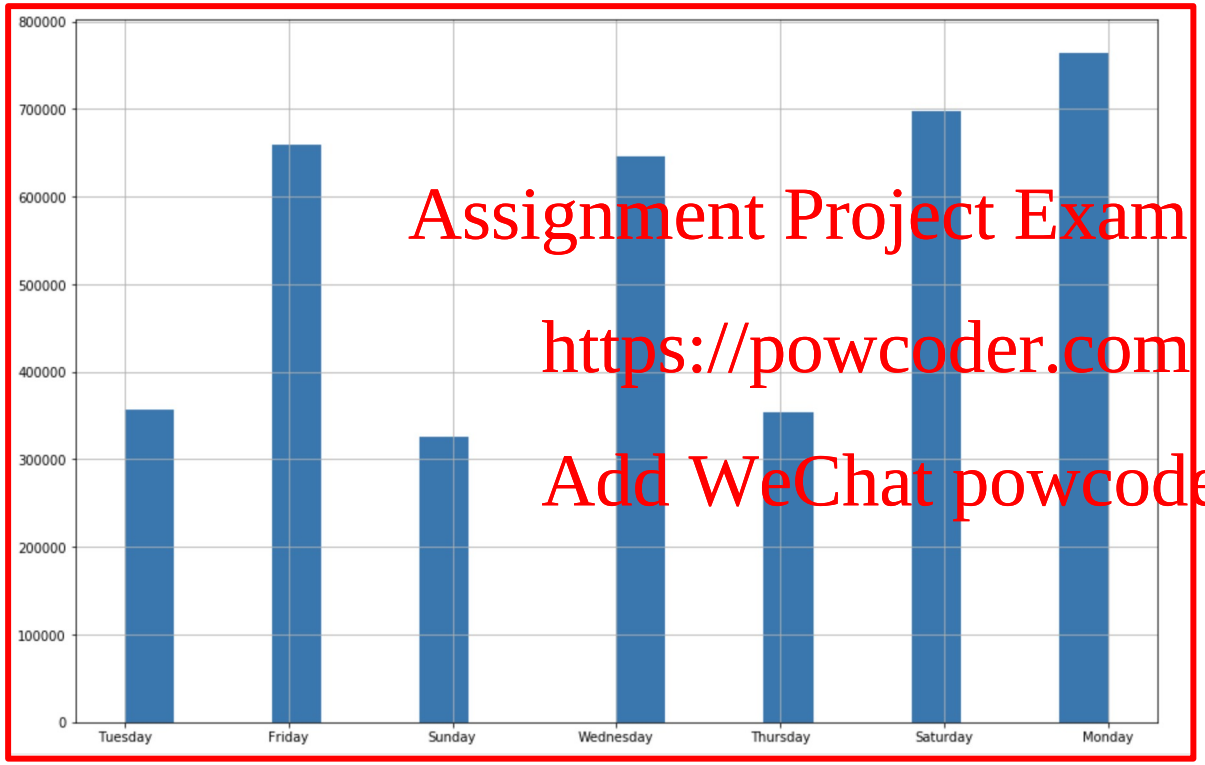
Comment: Use of multiple colors not relevant, title and axis descriptions are missing.

Indicative mark: 3/10

Rubric: Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

WEEKDAYS

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

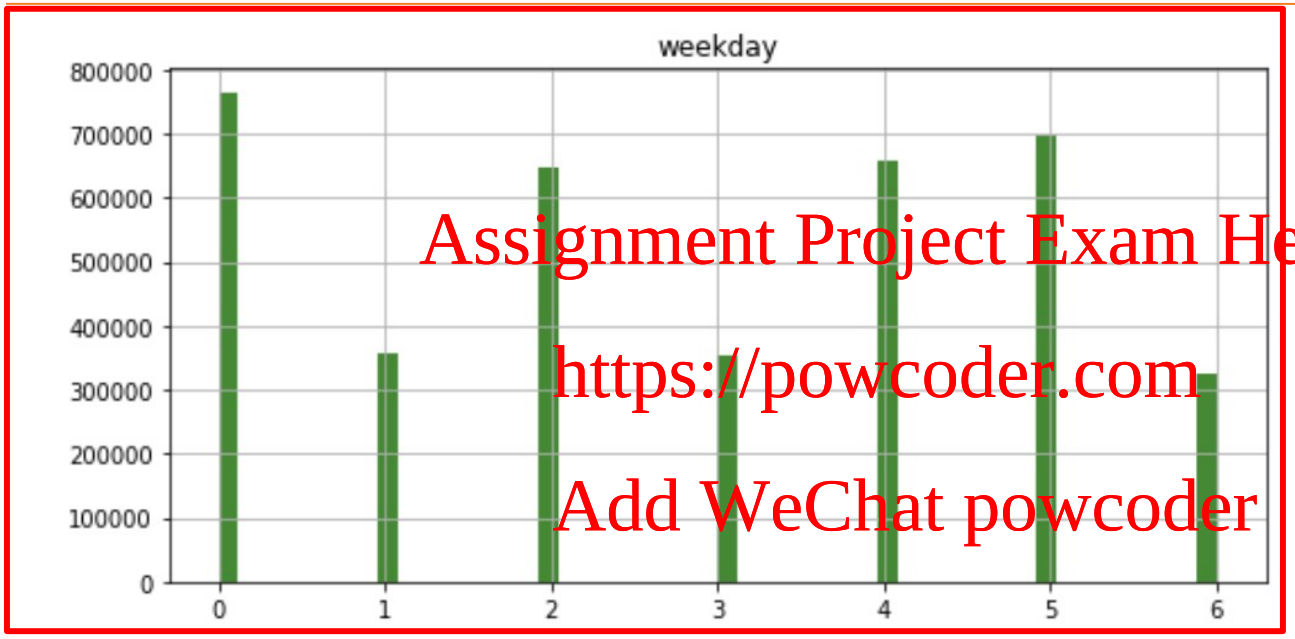
Comment: The chart lacks title and weekdays should be sorted.

Indicative mark: 5/10

Rubric: The report generally has a clear structure, although it would benefit from further development. Some errors and inconsistencies.

WEEKDAYS

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

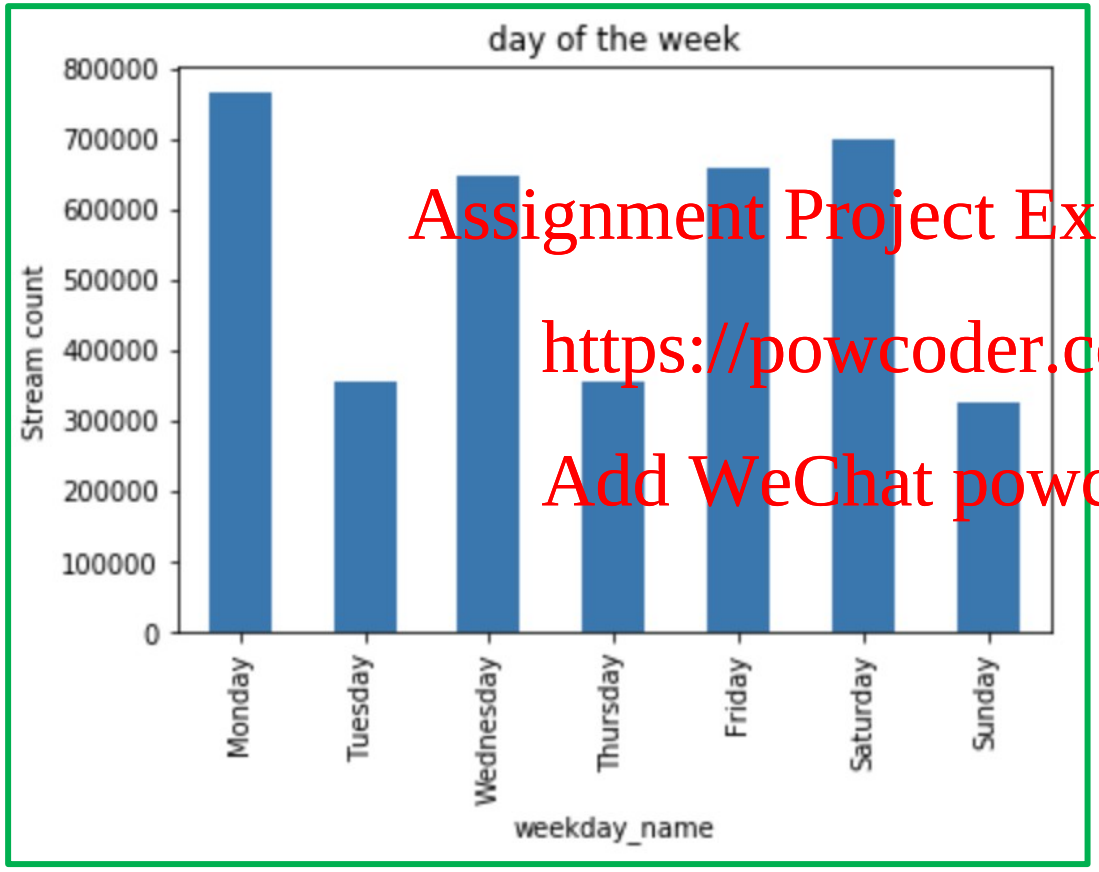
Comment: Counterintuitive labels on x axis.

Indicative mark: 5/10

Rubric: The report generally has a clear structure, although it would benefit from further development. Some errors and inconsistencies.

WEEKDAY

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

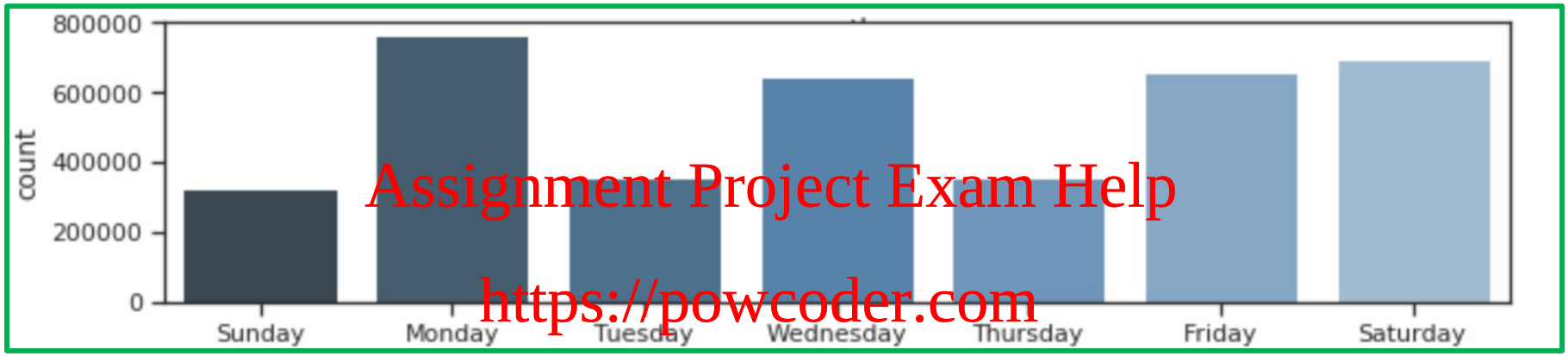
Comment: Well and logically presented at the basic level.

Indicative mark: 7/10

Rubric: Good solid work. Demonstrates effort to provide a usable document. Meets all of the criteria at the basic level.

WEEKDAY

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

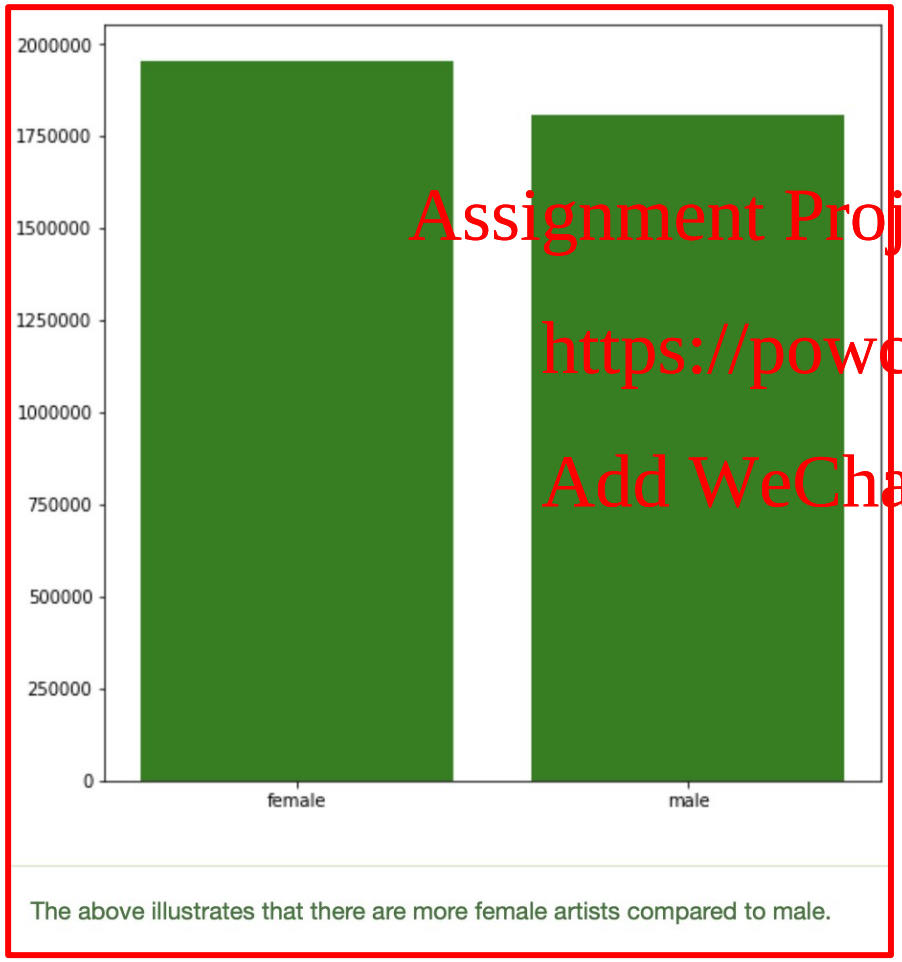
Comment: Well and logically presented at the basic level.

Indicative mark: 7/10

Rubric: Good solid work. Demonstrates effort to provide a usable document. Meets all of the criteria at the basic level.

GENDER

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

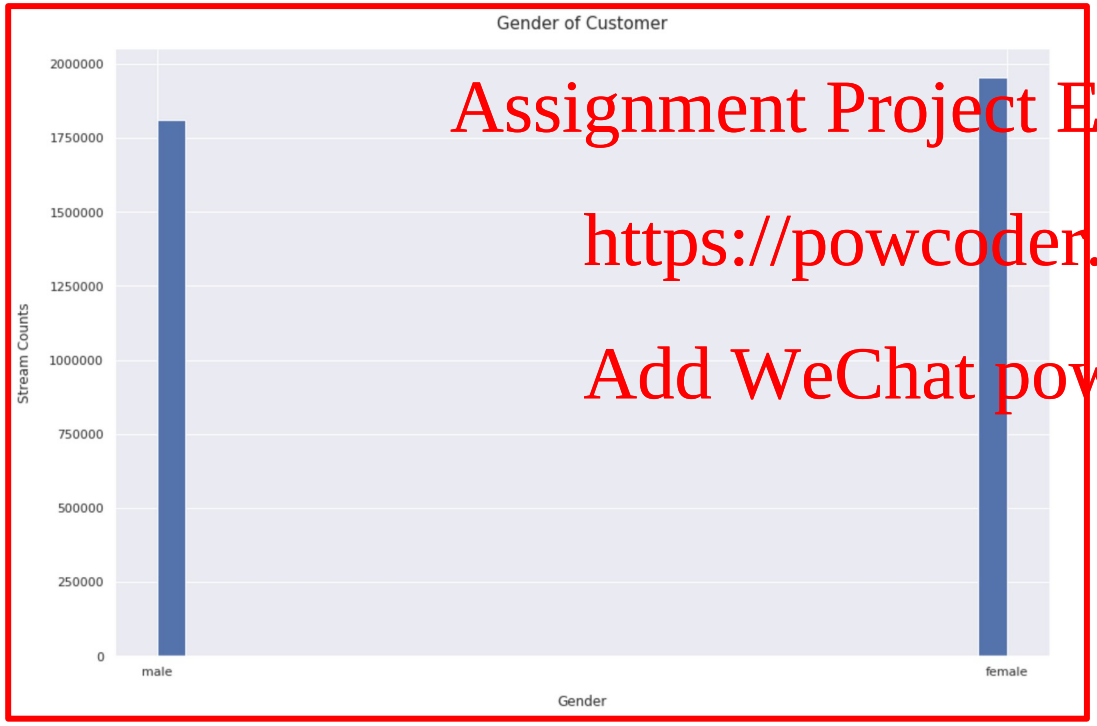
Comment: Poor visualization, hard to seen any difference, could be improved by using a different type of chart. This kind of narrative does not meet the basic requirements of the assignment.

Indicative mark: 4/10

Rubric: Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

GENDER

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

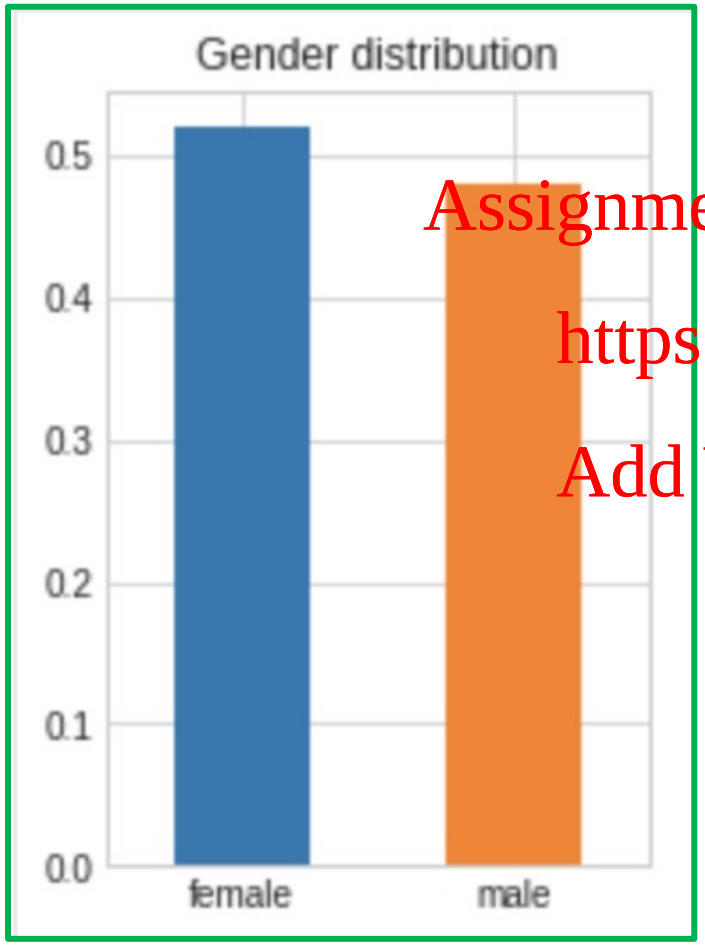
Comment: Poor visualization, hard to seen any difference, scale should be adjusted.

Indicative mark: 4/10

Rubric: Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

GENDER

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

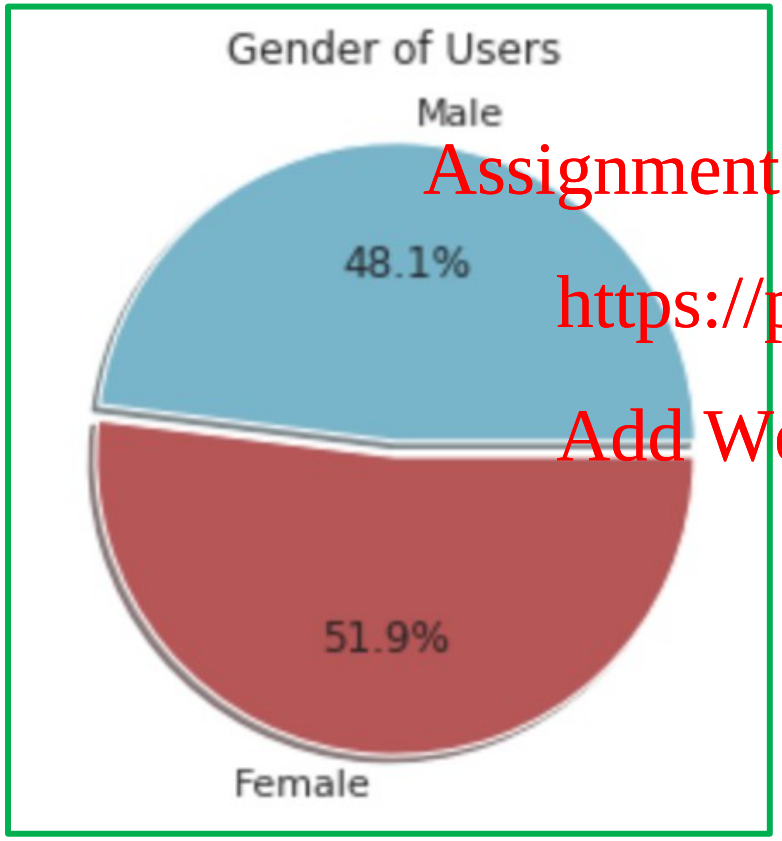
Comment: Well and logically presented.

Indicative mark: 7/10

Rubric: Good solid work. Demonstrates effort to provide a usable document. Meets all of the criteria at the basic level.

GENDER

UCL
SCHOOL OF
MANAGEMENT



Comment: Well and logically presented. Includes slice labels and labelling percentage.

Assignment Project Exam Help

Indicative mark: 8/10

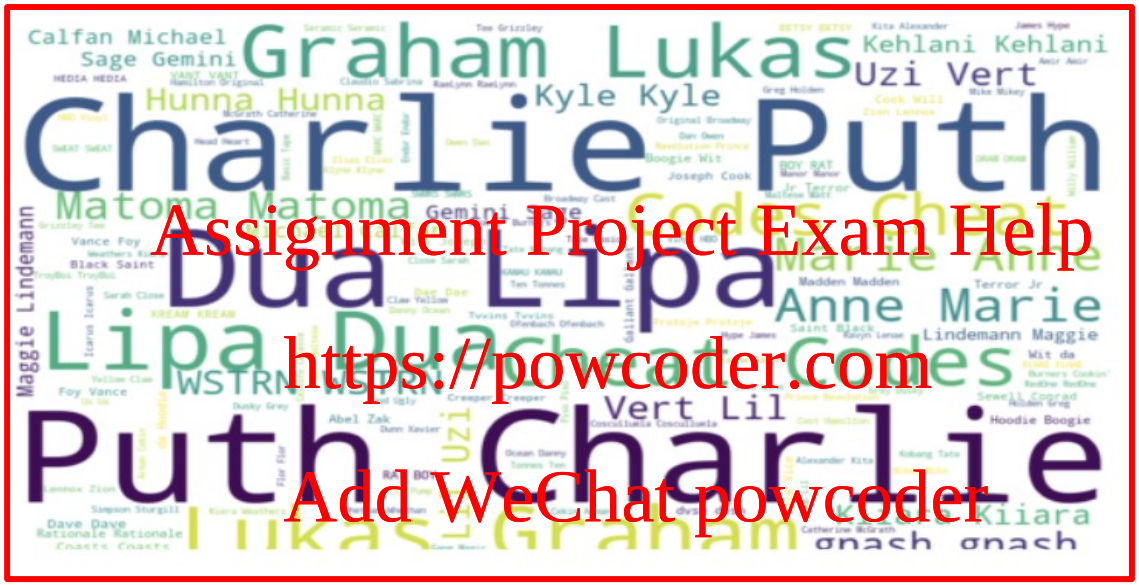
https://powcoder.com

Rubric: Professionally presented and well laid out. Meets all of the criteria at an advanced level.

Add WeChat powcoder

DIFFERENT ANGLE

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

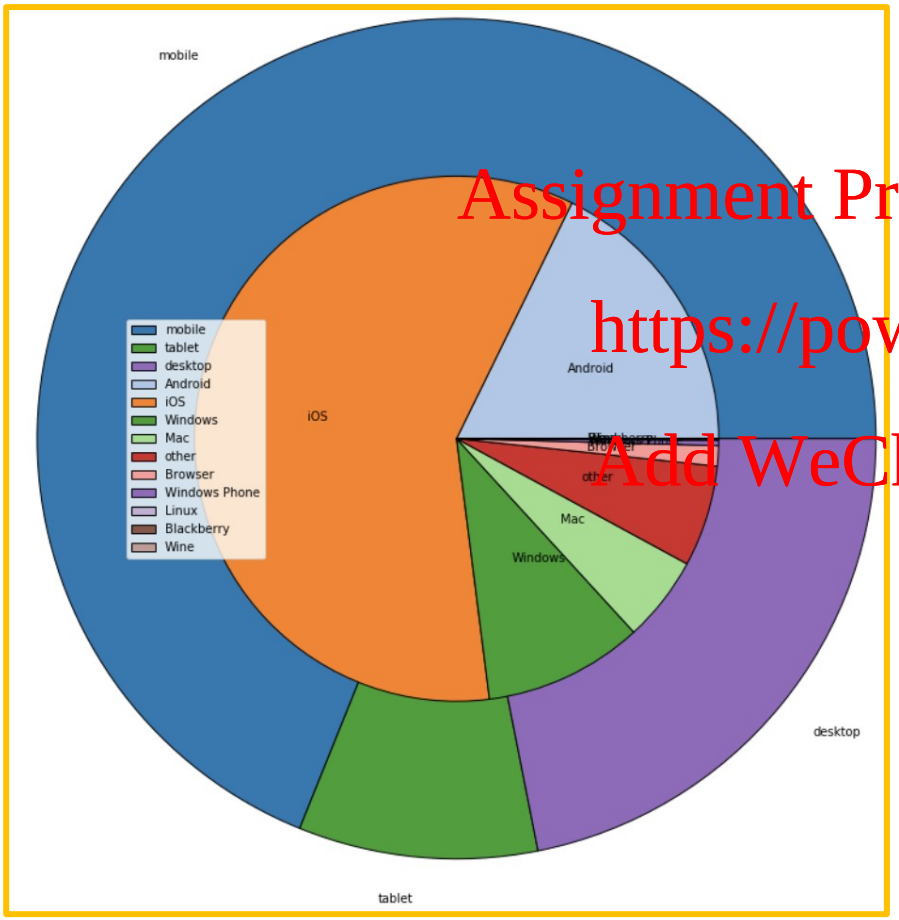
Comment: Poor visualization, does not show anything useful.

Indicative mark: 4/10

Rubric: Incoherent appearance / layout. This is usable in only a very few places, and there are a considerable number of errors and inconsistencies.

DIFFERENT ANGLE

UCL
SCHOOL OF
MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

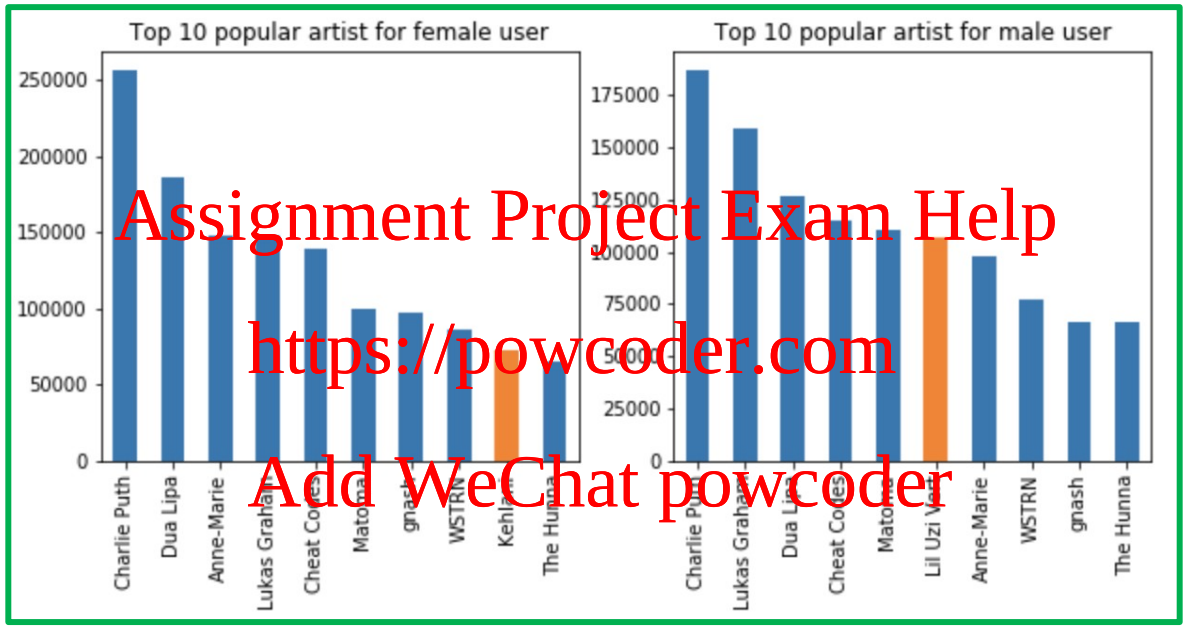
Comment: Interesting visualization, showing lots of information although would benefit from using distinct colors and binning some categories to make it clear, as well as adjusting the scale.

Indicative mark: 5/10

Rubric: The report generally has a clear structure, although it would benefit from further development. Some errors and inconsistencies.

DIFFERENT ANGLE

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

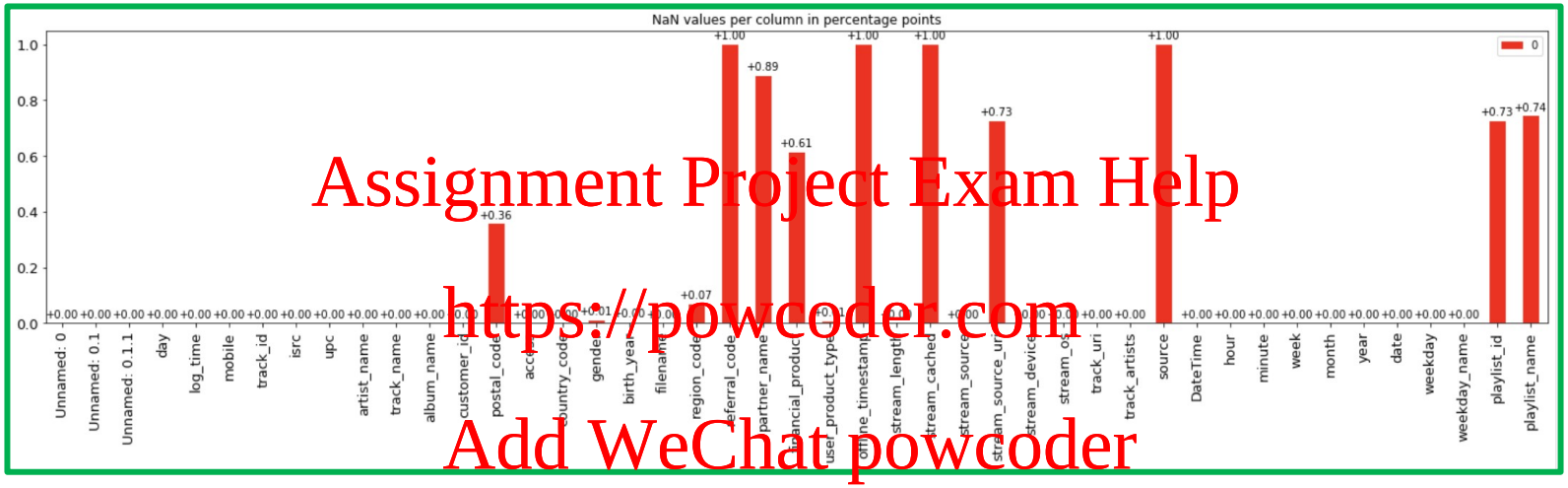
Comment: Very good use of color, clearly emphasizes the point.

Indicative mark: 7/10

Rubric: Good solid work. Demonstrates effort to provide a usable document. Meets all of the criteria at the basic level.

DIFFERENT ANGLE

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

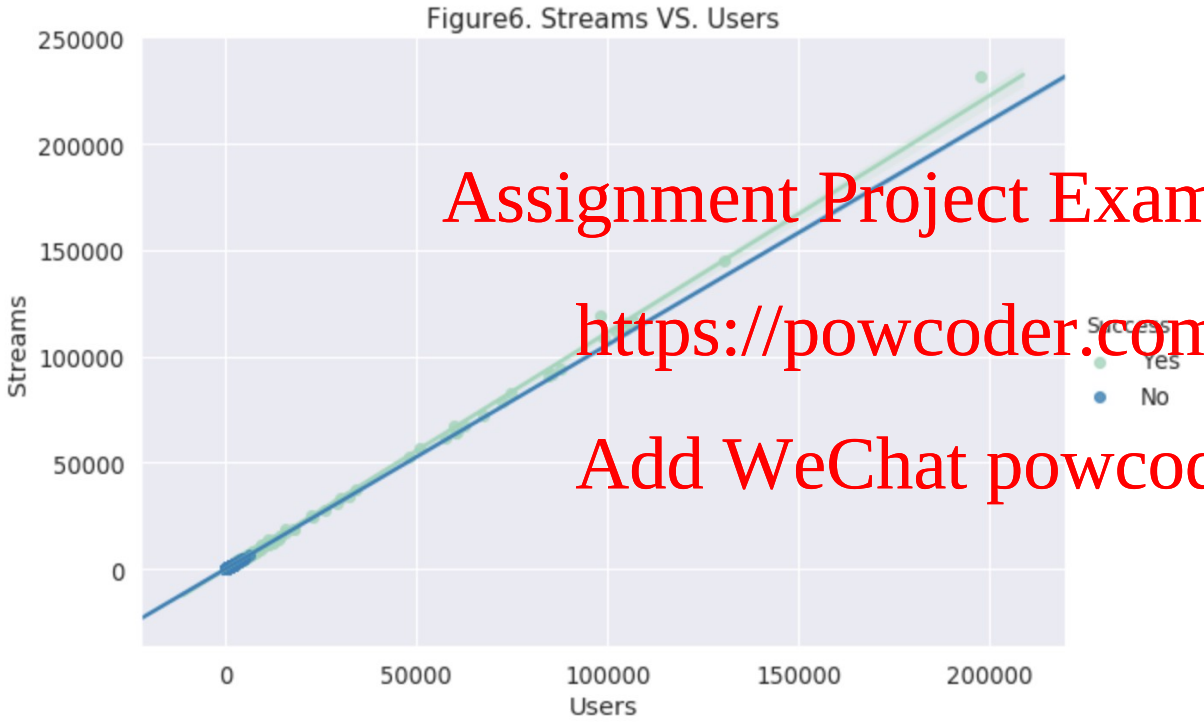
Comment: Very interesting way to show missing values, nicely presented, demonstrates effort, would benefit from removing the legend.

Indicative mark: 7/10

Rubric: Good solid work. Demonstrates effort to provide a usable document. Meets all of the criteria at the basic level.

DIFFERENT ANGLE

UCL SCHOOL OF MANAGEMENT



Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Well and logically presented, from an interesting angle with insightful narrative.

Indicative mark: 9/10

Rubric: Excellent work which demonstrates original thought and effort. The presentation enables communications well.

Random Lovers vs Loyal Fans

Are the successful artists better at attracting more random audience or loyal fans? It turns out they are good at both. The statistics and figure show that on average, the songs of the successful artists receive a larger group of audience. Meanwhile, the fitted line success = 'Yes' has a higher slope, this means that more users listen to their songs more than once.

NARRATIVE

UCL
SCHOOL OF
MANAGEMENT

2.5 Conclusion

This concludes our data cleaning and exploration.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: These comments are not relevant, therefore should not be included. The word limit for all 3 parts of the assignment is set at around 2000 words, such narratives are reducing the quality of the submission.

Indicative mark: 1/10

Rubric: It appears that it may have been misunderstood what was required. The points are either copied from the case with no analysis or not clearly related to the events in the case.

NARRATIVE

UCL SCHOOL OF MANAGEMENT

Artist name

From this chart above, we oberse **the top 10 artists are Charlie Puth, Dua Lipa, Lukas Graham, Cheat Codes and Anne-Marie, Matoma, gnash, WSTRN, Lil Uzi Vert, The Hunna.** Charlie Puth, particularly, has more than 400,000 stream counts and others are around 200,000.

Assignment Project Exam Help

https://powcoder.com

Comment: This narrative duplicates the information that shown in the visualization and does not include any insight nor reasoning, which are the key.

Add WeChat powcoder

Indicative mark: 2/10

Rubric: Significant errors / omissions / unclear points. Much more work is needed to develop and support assertions.

NARRATIVE

UCL
SCHOOL OF
MANAGEMENT

This shows the dataset covers 661 unique artists. Charlie Puth ranks top 1about the frequency of streams. The top 10 artists account for around 63% of the total data and they should be paid more attention.

Assignment Project Exam Help

https://powcoder.com

Comment: Very little insight is presented, reasoning is missing.

Add WeChat powcoder

Indicative mark: 5/10

Rubric: Arguments are generally along the right lines but a little weak in places. Scope to increase depth and breadth of analysis in many instances. Quite a few key points missed.

NARRATIVE

UCL SCHOOL OF MANAGEMENT

2.4.2 Artist name

Interesting findings:

1. The distribution of streams by artists is highly skewed. Most artists are streamed less than 44 times (median is 44), but the **max is 446783 by Charlie Puth** .
2. The graph of the count of streams by the top 30 artists shows an **exponential (decreasing) distribution**.

Assignment Project Exam Help

https://powcoder.com

Comment: No reasoning included, although all other requirements are satisfied. This narrative puts the visualization in context of the case study and is very relevant to the point of the assignment.

Add WeChat powcoder

Indicative mark: 6/10

Rubric: A solid analysis which covers many of the key issues, although scope for further analysis in some places. A few key points missed.

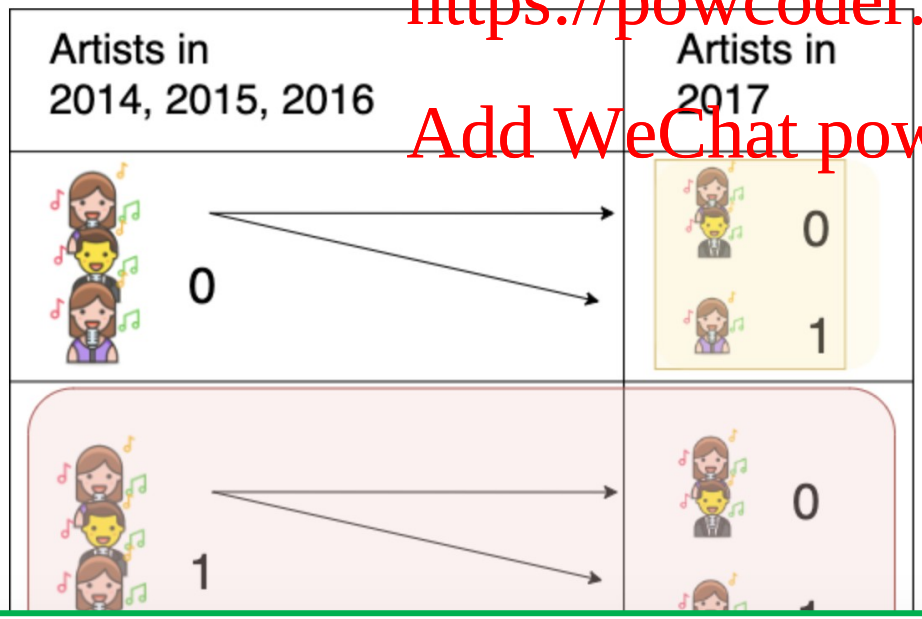
DEFINING THE PROBLEM

UCL SCHOOL OF MANAGEMENT

3.1 Forming our predictive problem

For our predictive problem we've decided to not deal with any artist who became succesful (i.e. featured on top 4 key playlists) in years 2014 to 2016 (red shaded area in the graph). In addition, we've selected to keep the information of who became succesful in 2017 while removing any transcactions from this year. The concept here is that based on transcactions from 2014 to 2016 we are trying to predict who became succesful in 2017 (yellow shaded area in the graph).

In the next code blocks, we create labels of who became succesful in 2017 as well as who became succesful in years 2014-2016. Based on these labels we subset our initial dataset.



Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: Unique and well-thought approach, not only explained but also presented with a figure.

Indicative mark: 8/10

Rubric: Very good points, with well-supported arguments which relate clearly to events in the case study.

METHODOLOGY

UCL SCHOOL OF MANAGEMENT

User-based Features

User-based features can be strong indicators for the success of artists. Based on the exploratory analysis before, we use age, gender, and **user-product** breakdown as the user-based features. In addition, previous analysis shows that successful artists has higher concentration of streams from May-June, so we add a **month-stream breakdown** here to capture this propensity.

Assignment Project Exam Help

https://powcoder.com

Comment: Logically presented, well reasoned, but lacking depth.

Add WeChat powcoder

Indicative mark: 7/10

Rubric: Points are generally relevant. But could go further in identifying the breadth of issues raised by the case. Overall good, solid work. Meets the requirements of the assessment.

METHODOLOGY

UCL SCHOOL OF MANAGEMENT

String handeling

Because some artists appear twice with the only difference beeing a capitalized letter, and we put huge emphasize on the success of individual artists, we need to standardize the spelling of artists' name. Thats why we lowercase every letter right at the beginning to avoid later problems.

To be noted is that the data also contains some artists with a "&" in their name, indicating more than one artist on these particular rows. Due to ambiguity in how to deal with those instances, it was decided to leave them as is, and count them as one artist, which makes sense as these instances are mostly duos like "zion & lennox". Compared to this some bands also contain more than one person and count as one artist, and features are accounted for seperately in the track name column.

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Important issue has been identified, handled, and explained.

Indicative mark: 8/10

Rubric: Very good identification of the key issues in the case. Meets the requirements at an advanced level.

METHODOLOGY

UCL SCHOOL OF MANAGEMENT

Playlist Score is calculated as follows:

$$ Playlist Score = \sum_{i=1}^{20} (nstreams_i \times passionscore_i \times weight_i) $$

Where variables are:

1. **n_streams** is the number of streams that playlist ith had overall
2. **passion_score** is the passion score of ith playlist (n_streams/unique_listeners)
3. **weight** is n times an artist has been featured on playlist ith/n times artist has been featured in all his/her top 20 playlists

Intuition:

The reason for this score is that it combines overall popularity/success/universality of a playlist (n_streams*passion_score) with the relative weight of playlists within top_20 playlists of a particular artist. Although this is a single number, this number holds a lot of information on it. We will later see that the playlist score actually has the biggest correlation with artist successfulness dependent variable.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

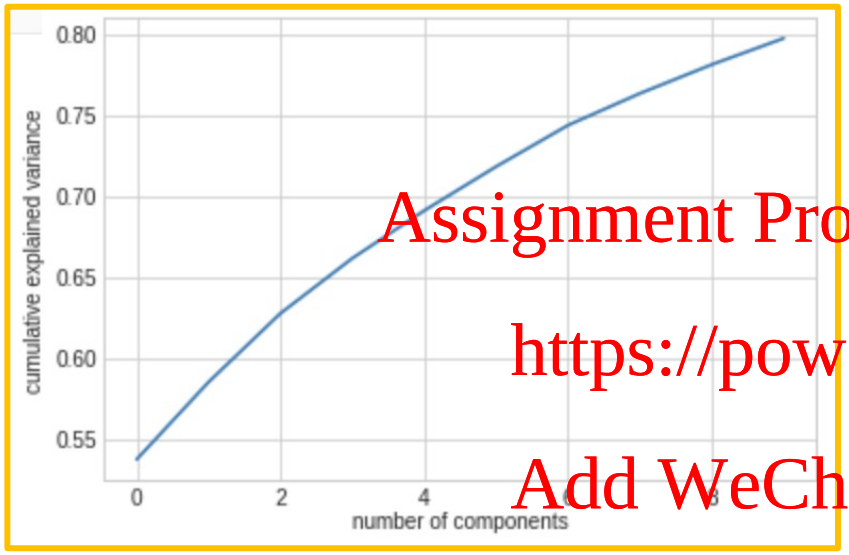
Comment: Going an extra mile in understanding the problem, demonstrates preparation to tackle the problem. Use of LaTeX to write an equation is appreciated.

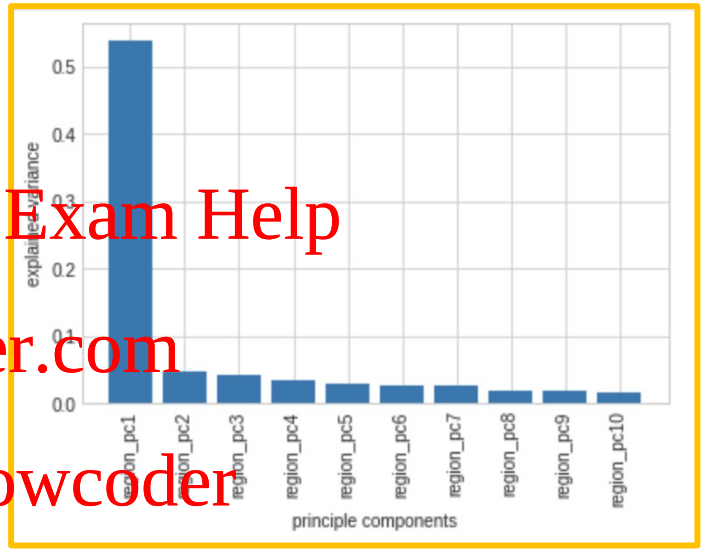
Indicative mark: 8/10

Rubric: Excellent work. Points are entirely relevant and focus clearly on the issues in the case. Very good identification of the key aspect that exceeds the scope of the basic requirements.

PCA

UCL SCHOOL OF MANAGEMENT





Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Clearly presented explained variance and/or cumulative variance. Any of the above sufficiently illustrates the concept.

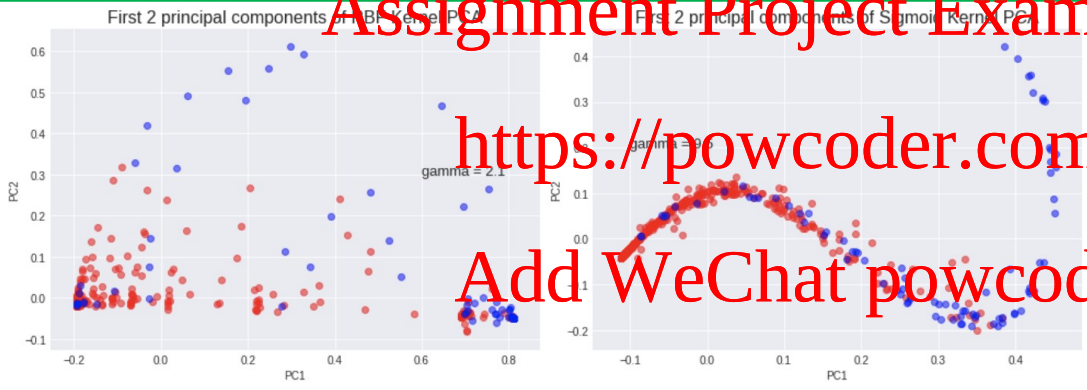
Indicative mark: 6/10

Rubric: Generally well-structured and solid in appearance and usability. Some minor errors and inconsistencies in places.

PCA

UCL SCHOOL OF MANAGEMENT

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder



RBF and Sigmoid Kernel plots

The first two components from the kernel PCA are plot against each other. The red and blue plot represents the not success and successful artists respectively. The right hand side of the Sigmoid kernel and the top part of the RBF kernel shows quite clear separation between the two groups of artists.

Comment: Presenting the relationship between the first two components shows understanding of the subject and beneficial for the overall analysis. Would benefit from an insightful comment for the purpose of assessing the narrative.

Indicative mark: 8/10

Rubric: Professionally presented and well laid out. Meets all of the criteria at an advanced level.

CODE ANNOTATION

UCL SCHOOL OF MANAGEMENT

```
data['age']=2020-data['birth_year']

histogram_birth_year = data['age']
histogram_birth_year = histogram_birth_year.hist(figsize=(10, 7), bins=300,
                                   xlabelsize=12, ylabelsize=12,
                                   color="lightgreen", grid=False),plt.title("Customer age repartition"),plt.axis([10,
                                                                                                                  )
```

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Basic code, no comments, not using variables in the appropriate way.

Indicative mark: 4/10

Rubric: Some evidence of additional sources, although not all are very robust

CODE ANNOTATION

UCL
SCHOOL OF
MANAGEMENT

```
# Taking a closer look at 1950-2017 and noticing the peaks at 1980, 1990
df2[df2.birth_year >= 1950].groupby('birth_year')['birth_year'].count().plot(kind='bar', figsize=(20,5));
```

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Nice code, with basic comments.

Indicative mark: 6/10

Rubric: Good sources in places but use of quality sources could be extended.

CODE ANNOTATION

UCL SCHOOL OF MANAGEMENT

```
# histogram age distribution
sns.distplot(sub_data['age'], hist=True, kde=True,
             bins=50)

# setting x limits
plt.xlim(-1, 85)

#formatting
plt.xlabel('Age', fontsize = 20, labelpad=15)
plt.ylabel('Density', fontsize = 20, labelpad=15)
plt.title('Age Distribution of UK Spotify Users', fontsize=25).set_position([.5, 1.07])
plt.tick_params(axis='both', which='major', pad=10)
plt.grid(linestyle='')
plt.gcf().set_size_inches(25,14.5)
sns.despine()
plt.show()
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: Nice code, comments enable reader to easily understand the code.

Indicative mark: 8/10

Rubric: Very good work with relevant and robust sources. Very good referencing and documentation of code.

CODE

UCL
SCHOOL OF
MANAGEMENT

```
# Plot differently through an heatmap
import seaborn as sns

map = region_dataframe
plt.figure(figsize=(15,10))
sns.heatmap(map,cmap='twilight')
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: Using a built-in function's name as a variable will make it difficult to use the function later in the notebook and is generally frowned upon.

Indicative mark: 1/10

Rubric: Very little or no evidence of use of sources beyond the course materials

CODE

UCL
SCHOOL OF
MANAGEMENT

```
# use defined function to check
vif_cal(input_data=final_df1, dependent_col=final_df1["1"])
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
<ipython-input-193-d1e123f5cc39> in <module>
      1 # use defined function to check
----> 2 vif_cal(input_data=final_df1, dependent_col=final_df1["1"])

<ipython-input-191-a17e75ca48c2> in vif_cal(input_data, dependent_col)
      2
      3 def vif_cal(input_data, dependent_col):
----> 4     x_vars=input_data.drop([dependent_col], axis=1)
      5     xvar_names=x_vars.columns
      6     for i in range(0,xvar_names.shape[0]):

//anaconda3/envs/py37/lib/python3.7/site-packages/pandas/core/frame.py in drop(self, labels, axis, index, co
lumns, level, inplace, errors)
   4115             level=level,
   4116             inplace=inplace,
-> 4117             errors=errors,
   4118         )
   4119
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: Leaving errors shows lack of understanding of the key concepts of coding and lack of problem-solving attitude which is crucial for a successful analysis.

Indicative mark: 1/10

Rubric: Very little or no evidence of use of sources beyond the course materials

CODE

UCL SCHOOL OF MANAGEMENT

```
#top20 playlists decided by # unique users/customer_ids per playlist

top20_b=data.groupby(["artist_name",'playlist_name'])["customer_id"].nunique().reset_index(name="# of users")

c = top20_b.groupby(["artist_name"]).apply(lambda x: x.sort_values(["# of users"], ascending = False)).reset_index(d

d = c.groupby('artist_name').head(20)

d.loc[d['artist_name'] == 'Anne-Marie']
```

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder

Comment: Using variable names that are meaningful will enable the reader (as well as the author) to understand the code more easily and will reduce the need for comments.

Indicative mark: 3/10

Rubric: Some limited evidence of additional sources, although not very robust ones

CODE

UCL
SCHOOL OF
MANAGEMENT

```
# Number of users per artist
def numberofusers_per_artist(data):
    unique_users = data.groupby('artist_name')['customer_id'].unique() #get unique values for customer_id for each artist
    unique_users = pd.DataFrame(unique_users)
    unique_users['unique_users'] = [len(i) for i in unique_users.customer_id] #return number of unique users for each artist
    unique_users.drop('customer_id', axis=1, inplace=True) #drops the customer_id column
    return unique_users

numberofusers_per_artist(all_artists)
```

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

Comment: Using functions is an important part of coding, so is using relevant comments.

Indicative mark: 7/10

Rubric: Good, solid sources and referencing. In places, the use of code source and documentation could be improved / extended but overall good, solid work.

THANK YOU

UCL
SCHOOL OF
MANAGEMENT

Assignment Project Exam Help

https://powcoder.com

Add WeChat powcoder