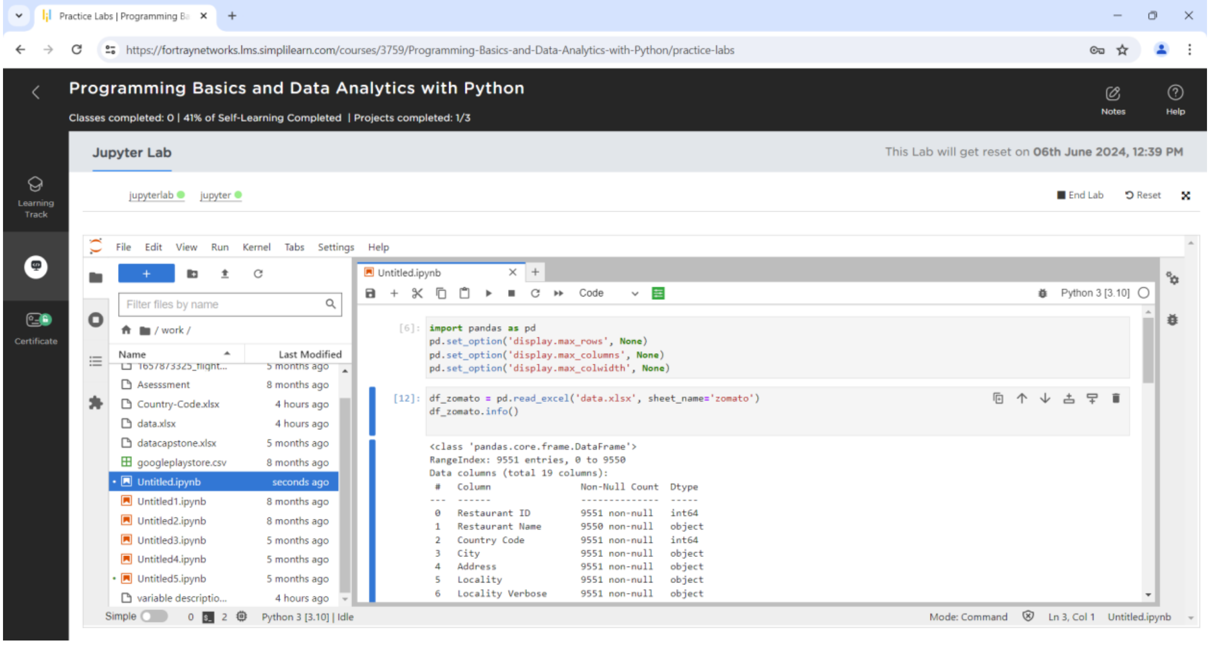
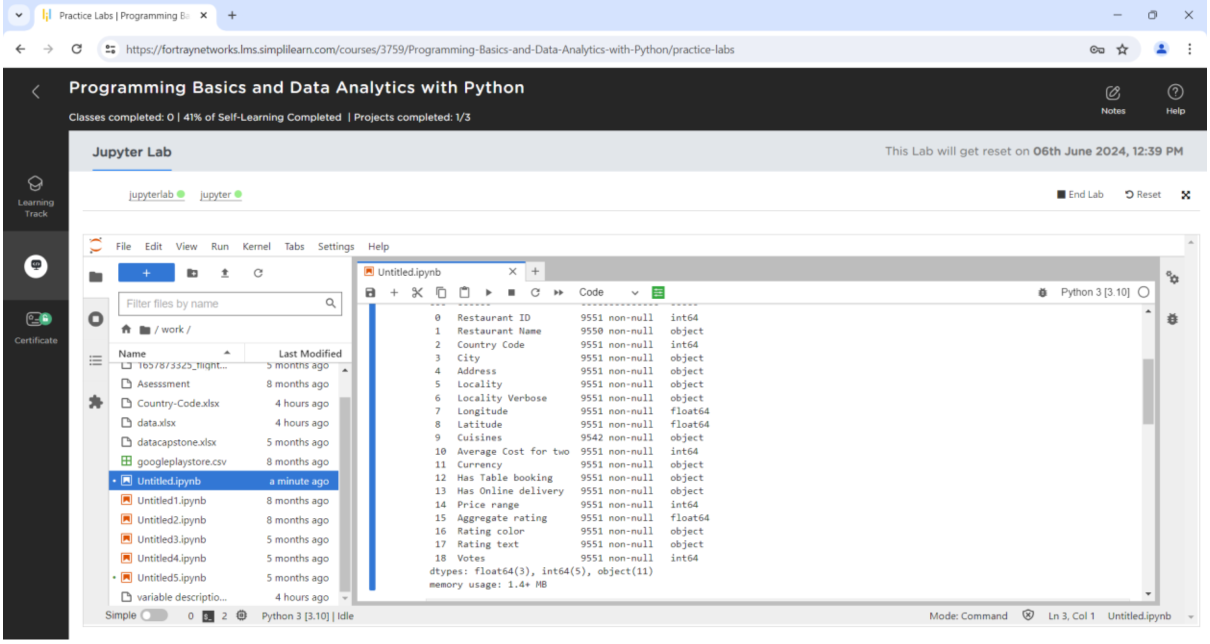
Data Preliminary analysis:

Perform preliminary data inspection and report the findings as the structure of the data, missing values,

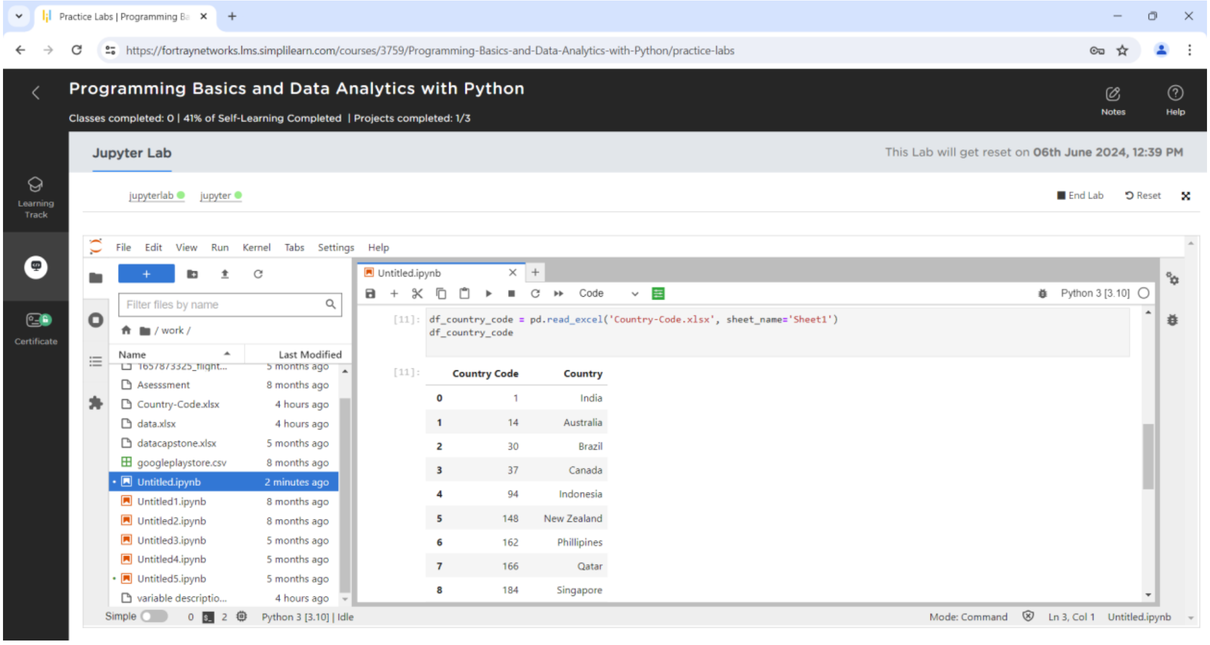
duplicates cleaning variable names etc.


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
Python 3 [3.10]
Mode: Command
Ln 3, Col 1

```
[6]: import pandas as pd
     pd.set_option('display.max_rows', None)
     pd.set_option('display.max_columns', None)
     pd.set_option('display.max_colwidth', None)

[12]: df_zomato = pd.read_excel('data.xlsx', sheet_name='zomato')
      df_zomato.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9551 entries, 0 to 9550
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Restaurant ID       9551 non-null   int64
 1   Restaurant Name     9550 non-null   object
 2   Country Code        9551 non-null   int64
 3   City                9551 non-null   object
 4   Address             9551 non-null   object
 5   Locality            9551 non-null   object
 6   Locality Verbose    9551 non-null   object
```


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
Python 3 [3.10]
Mode: Command
Ln 3, Col 1

```
 0   Restaurant ID       9551 non-null   int64
 1   Restaurant Name     9550 non-null   object
 2   Country Code        9551 non-null   int64
 3   City                9551 non-null   object
 4   Address             9551 non-null   object
 5   Locality            9551 non-null   object
 6   Locality Verbose    9551 non-null   object
 7   Longitude           9551 non-null   float64
 8   Latitude            9551 non-null   float64
 9   Cuisines            9542 non-null   object
 10  Average Cost for two  9551 non-null   int64
 11  Currency            9551 non-null   object
 12  Has Table booking   9551 non-null   object
 13  Has Online delivery  9551 non-null   object
 14  Price range         9551 non-null   int64
 15  Aggregate rating    9551 non-null   float64
 16  Rating color        9551 non-null   object
 17  Rating text         9551 non-null   object
 18  Votes               9551 non-null   int64
dtypes: float64(3), int64(5), object(11)
memory usage: 1.4+ MB
```



Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

This Lab will get reset on 06th June 2024, 12:39 PM

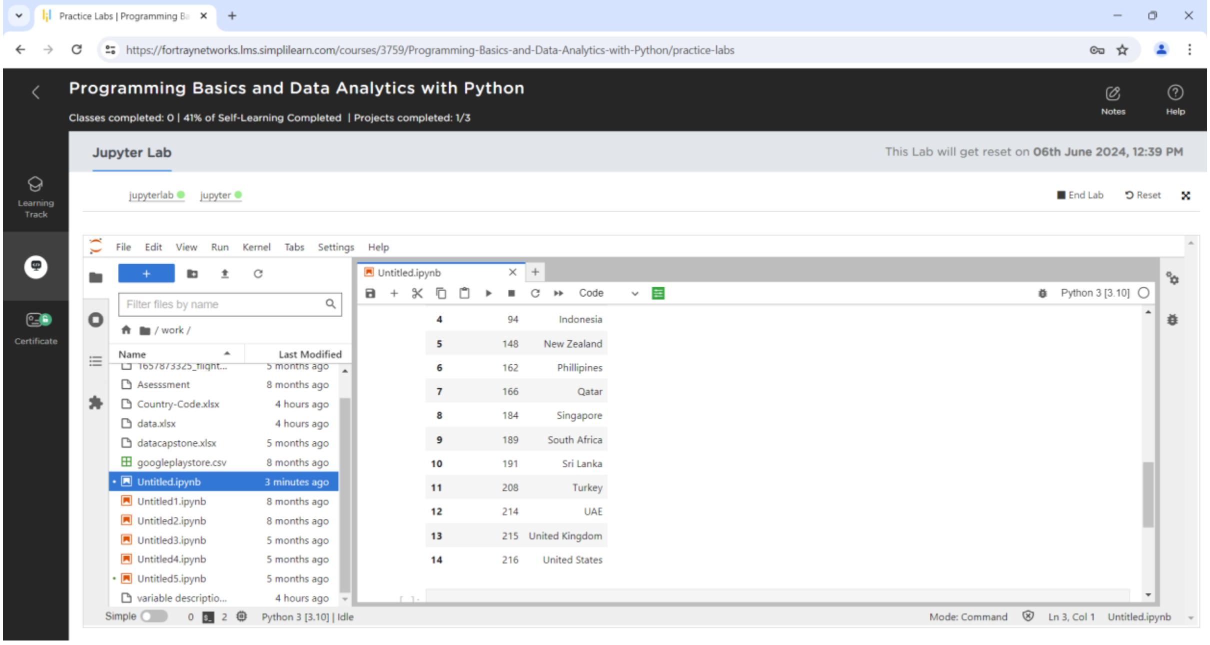
/ work /

Name	Last Modified
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	4 hours ago
data.xlsx	4 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	2 minutes ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago
Untitled5.ipynb	5 months ago
variable descriptio...	4 hours ago

```
[11]: df_country_code = pd.read_excel('Country-Code.xlsx', sheet_name='Sheet1')
      df_country_code
```

```
[11]:    Country Code      Country
      0             1        India
      1            14    Australia
      2            30       Brazil
      3            37       Canada
      4            94    Indonesia
      5           148  New Zealand
      6           162   Phillipines
      7           166        Qatar
      8           184    Singapore
```

0 2 Python 3 [3.10] | Idle



Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

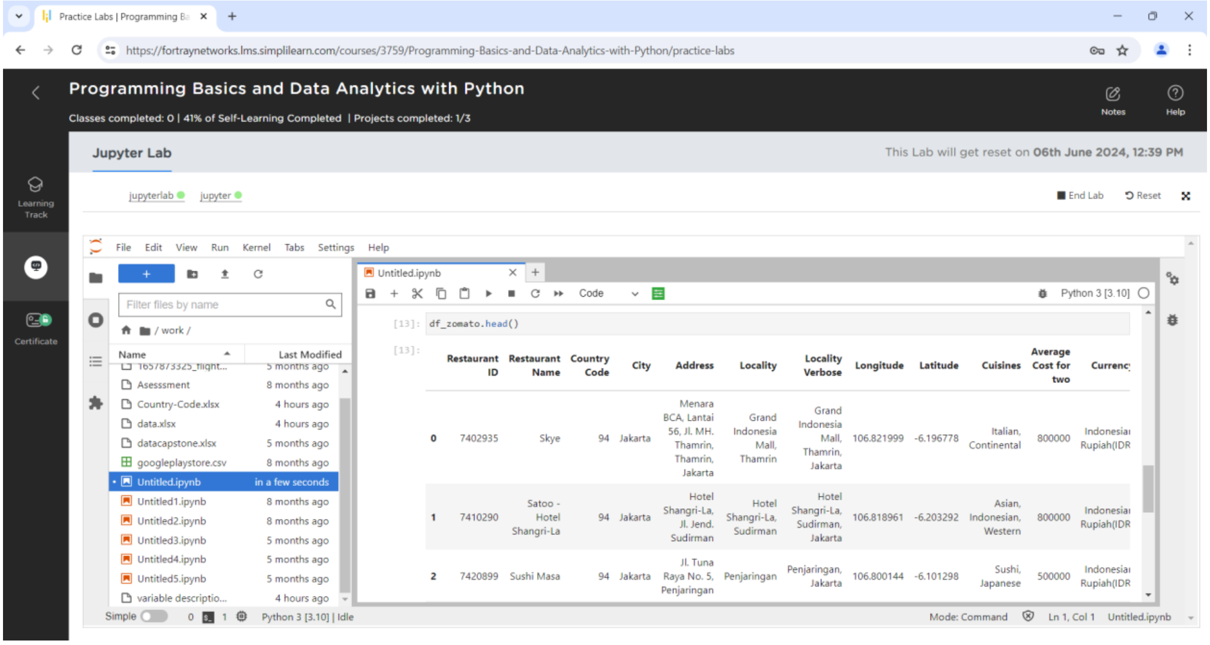
This Lab will get reset on 06th June 2024, 12:39 PM

/ work /

Name	Last Modified
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	4 hours ago
data.xlsx	4 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	3 minutes ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago
Untitled5.ipynb	5 months ago
variable descriptio...	4 hours ago

```
       4            94       Indonesia
       5           148     New Zealand
       6           162      Phillipines
       7           166           Qatar
       8           184       Singapore
       9           189    South Africa
      10           191       Sri Lanka
      11           208          Turkey
      12           214             UAE
      13           215  United Kingdom
      14           216   United States
```

0 2 Python 3 [3.10] | Idle

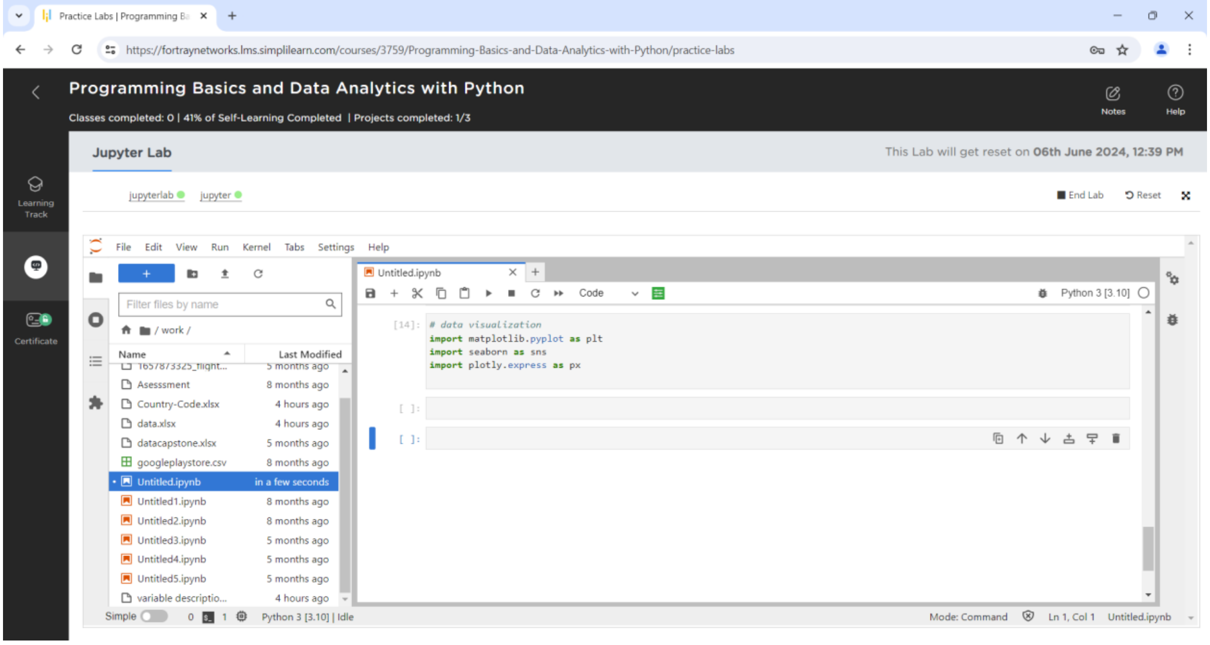

Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
Python 3 [3.10]
Mode: Command
Ln 1, Col 1

```
[13]: df_zomato.head()
```

[13]:

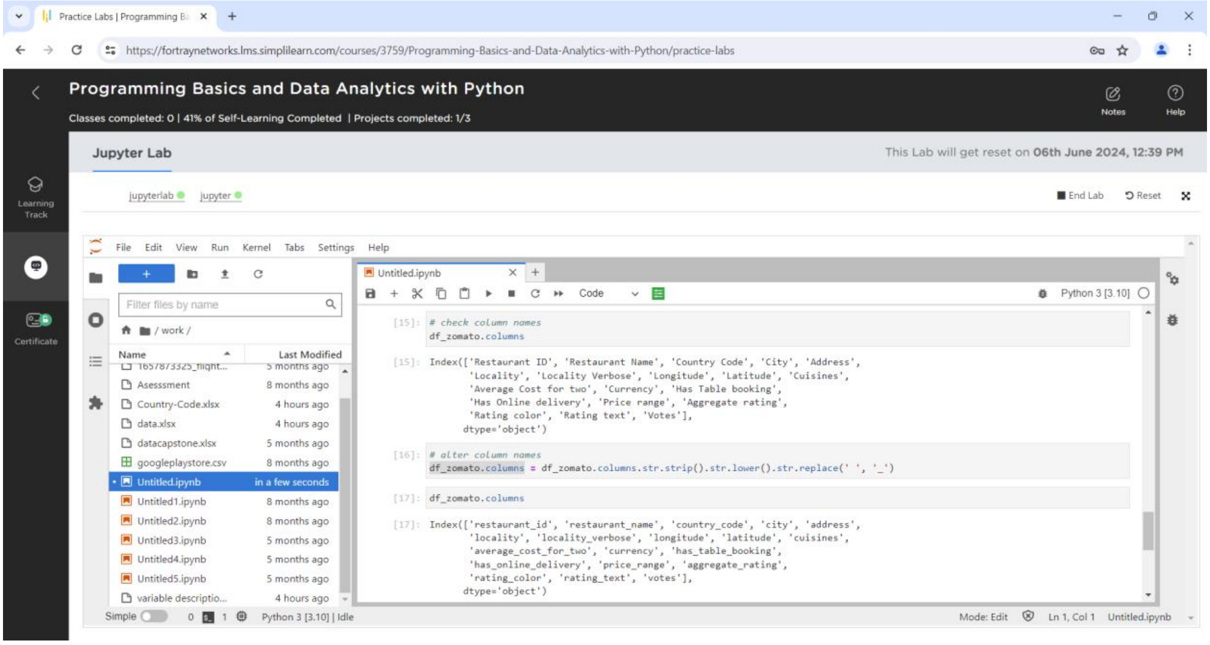
	Restaurant ID	Restaurant Name	Country Code	City	Address	Locality	Locality Verbose	Longitude	Latitude	Cuisines	Average Cost for two	Currenc
0	7402935	Skye	94	Jakarta	Menara BCA, Lantai 56, Jl. MH. Thamrin, Thamrin, Jakarta	Grand Indonesia Mall, Thamrin	Grand Indonesia Mall, Thamrin, Jakarta	106.821999	-6.196778	Italian, Continental	800000	Indonesia Rupiah(IDR
1	7410290	Satoo - Hotel Shangri-La	94	Jakarta	Hotel Shangri-La, Jl. Jend. Sudirman	Hotel Shangri-La, Sudirman	Hotel Shangri-La, Sudirman, Jakarta	106.818961	-6.203292	Asian, Indonesian, Western	800000	Indonesia Rupiah(IDR
2	7420899	Sushi Masa	94	Jakarta	Jl. Tuna Raya No. 5, Penjaringan	Penjaringan	Penjaringan, Jakarta	106.800144	-6.101298	Sushi, Japanese	500000	Indonesia Rupiah(IDR

Import packages for data visualization:


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
Python 3 [3.10]
Mode: Command
Ln 1, Col 1

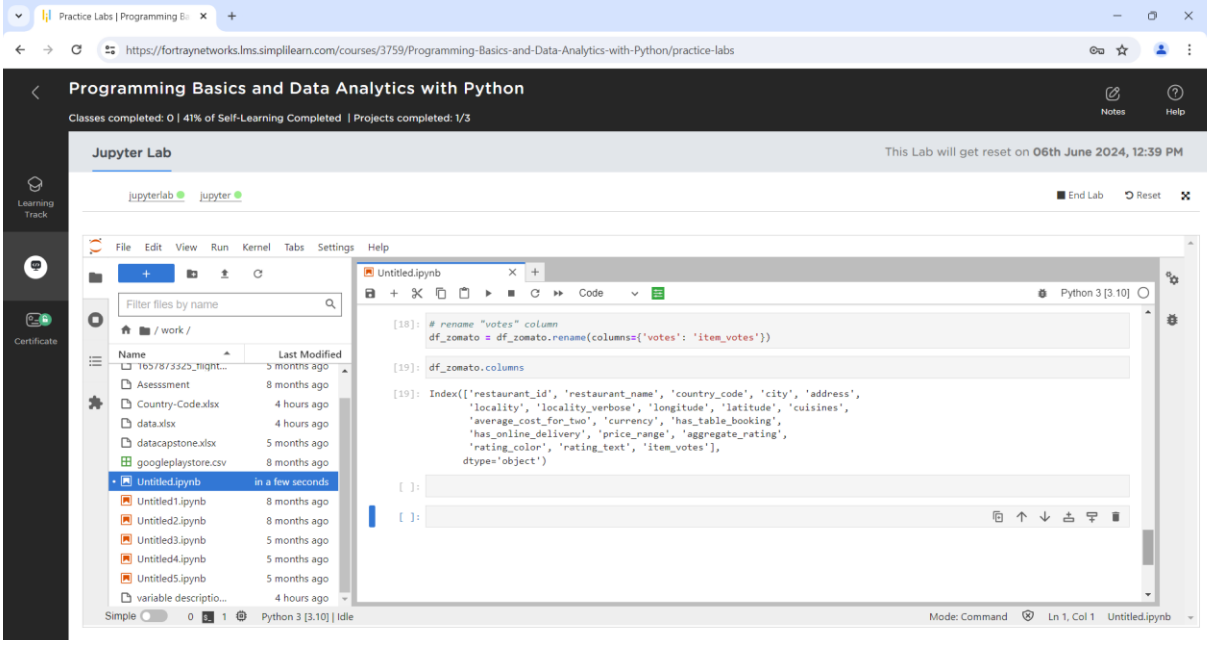
```
[14]: # data visualization
      import matplotlib.pyplot as plt
      import seaborn as sns
      import plotly.express as px

[ ]:

[ ]:
```

Convert all column names to lowercase and use underscores to separate words.


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Learning Track
Certificate
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python 3 [3.10]
Filter files by name
/ work /
Name
Last Modified
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx 4 hours ago
data.xlsx 4 hours ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb in a few seconds
Untitled1.ipynb 8 months ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled4.ipynb 5 months ago
Untitled5.ipynb 5 months ago
variable descriptio... 4 hours ago
Simple
Python 3 [3.10] | Idle
Mode: Edit
Ln 1, Col 1
Untitled.ipynb

```
[15]: # check column names
      df_zomato.columns

[15]: Index(['Restaurant ID', 'Restaurant Name', 'Country Code', 'City', 'Address',
             'Locality', 'Locality Verbose', 'Longitude', 'Latitude', 'Cuisines',
             'Average Cost for two', 'Currency', 'Has Table booking',
             'Has Online delivery', 'Price range', 'Aggregate rating',
             'Rating color', 'Rating text', 'Votes'],
            dtype='object')

[16]: # alter column names
      df_zomato.columns = df_zomato.columns.str.strip().str.lower().str.replace(' ', '_')

[17]: df_zomato.columns

[17]: Index(['restaurant_id', 'restaurant_name', 'country_code', 'city', 'address',
             'locality', 'locality_verbose', 'longitude', 'latitude', 'cuisines',
             'average_cost_for_two', 'currency', 'has_table_booking',
             'has_online_delivery', 'price_range', 'aggregate_rating',
             'rating_color', 'rating_text', 'votes'],
            dtype='object')
```

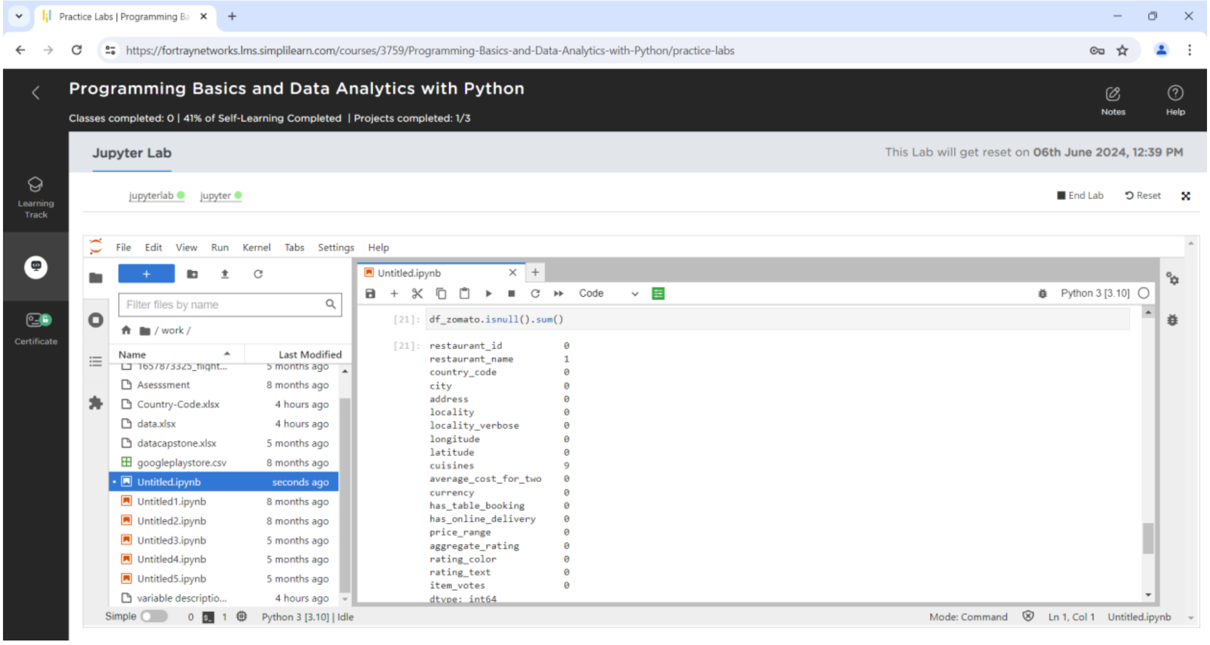
Rename the “votes” column to “item_votes:


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Learning Track
Certificate
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Untitled.ipynb
Code
Python 3 [3.10]
Filter files by name
/ work /
Name
Last Modified
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx 4 hours ago
data.xlsx 4 hours ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb in a few seconds
Untitled1.ipynb 8 months ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled4.ipynb 5 months ago
Untitled5.ipynb 5 months ago
variable descriptio... 4 hours ago
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled.ipynb

```
[18]: # rename "votes" column
      df_zomato = df_zomato.rename(columns={'votes': 'item_votes'})

[19]: df_zomato.columns

[19]: Index(['restaurant_id', 'restaurant_name', 'country_code', 'city', 'address',
             'locality', 'locality_verbose', 'longitude', 'latitude', 'cuisines',
             'average_cost_for_two', 'currency', 'has_table_booking',
             'has_online_delivery', 'price_range', 'aggregate_rating',
             'rating_color', 'rating_text', 'item_votes'],
            dtype='object')

[ ]:

[ ]:
```

Check null values:

Looks like there are some null values in the columns restaurant_name and cuisines

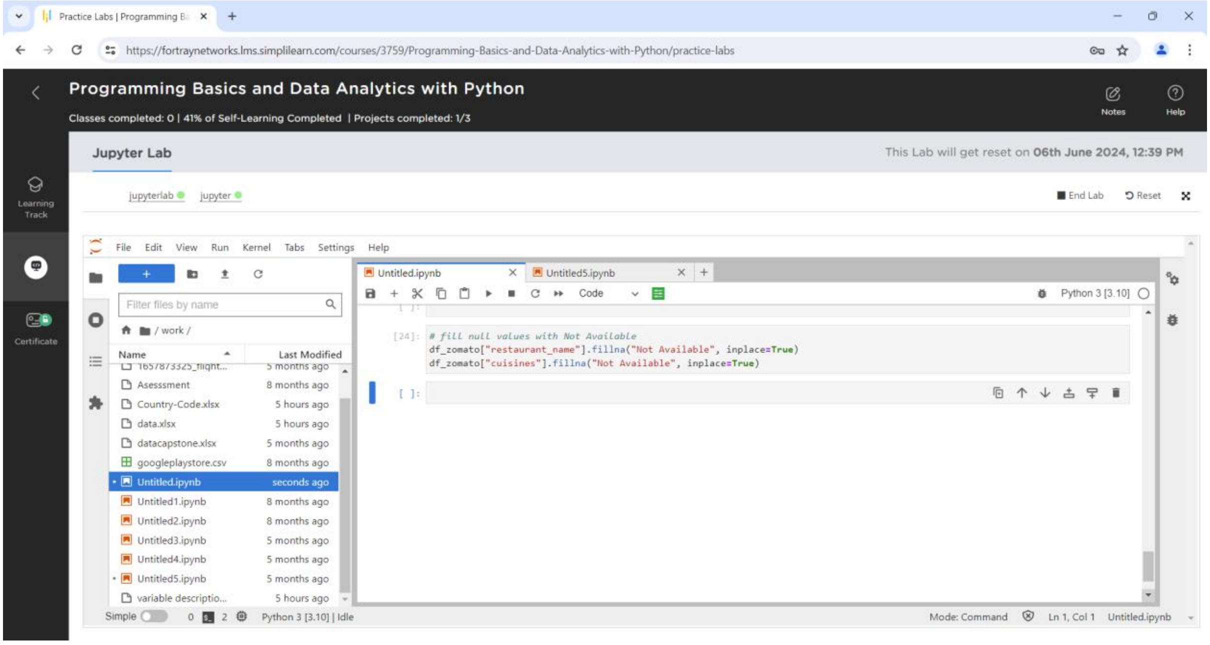

Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
End Lab
Reset
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Filter files by name
/ work /
Untitled.ipynb
Code
Python 3 [3.10]
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled.ipynb

Name	Last Modified
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	4 hours ago
data.xlsx	4 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago
Untitled5.ipynb	5 months ago
variable descriptio...	4 hours ago

```
[21]: df_zomato.isnull().sum()
```

```
[21]: restaurant_id           0
      restaurant_name         1
      country_code            0
      city                    0
      address                 0
      locality                0
      locality_verbose        0
      longitude               0
      latitude                0
      cuisines                9
      average_cost_for_two    0
      currency                0
      has_table_booking       0
      has_online_delivery     0
      price_range             0
      aggregate_rating        0
      rating_color            0
      rating_text             0
      item_votes              0
      dtype: int64
```

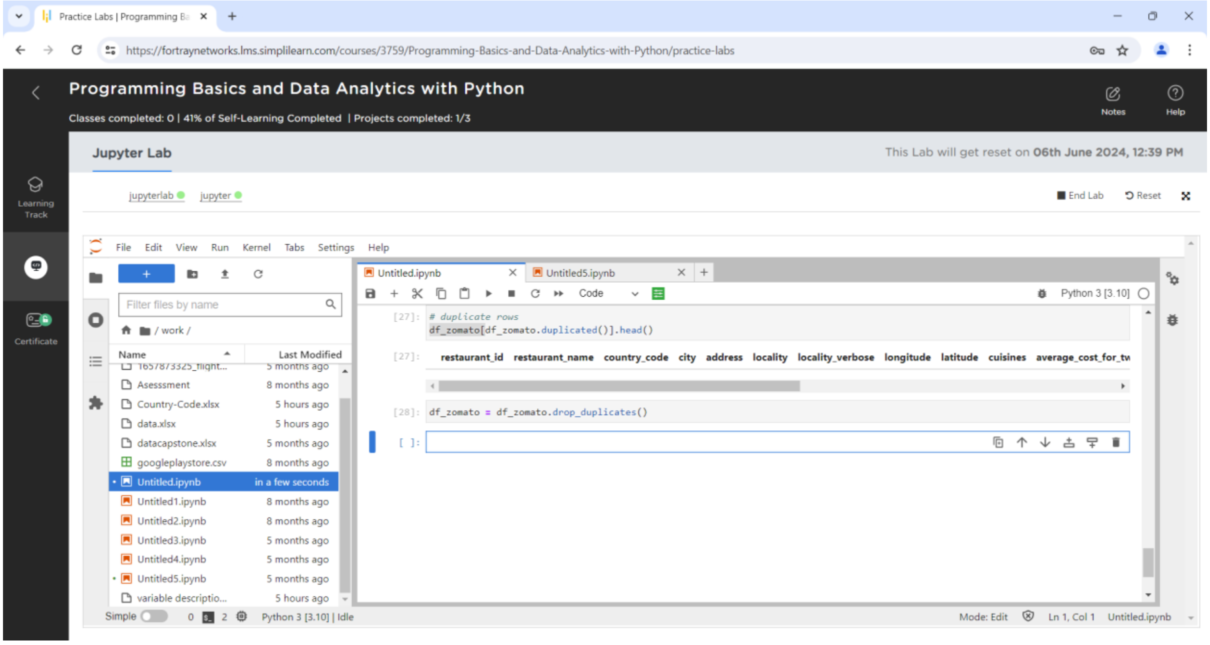
Fill the null values values in the columns restaurant_name and cuisines with "Not Available"


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
End Lab
Reset
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Filter files by name
/ work /
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled.ipynb

Name	Last Modified
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	5 hours ago
data.xlsx	5 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago
Untitled5.ipynb	5 months ago
variable descriptio...	5 hours ago

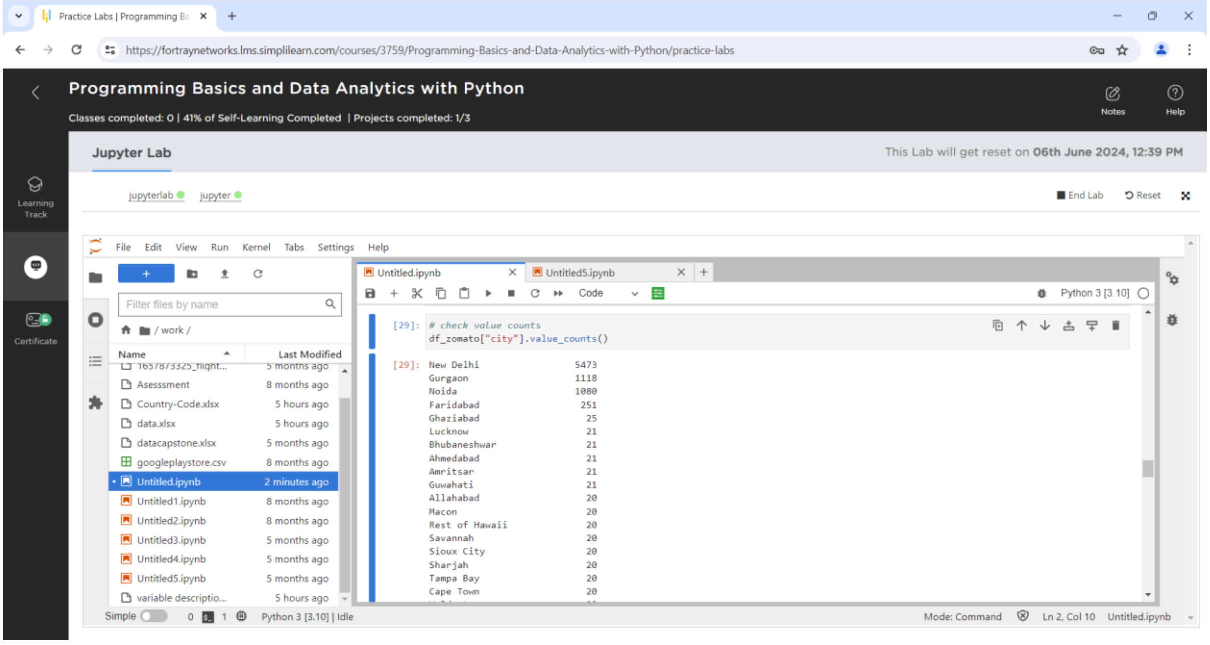
```
[24]: # fill null values with Not Available
      df_zomato["restaurant_name"].fillna("Not Available", inplace=True)
      df_zomato["cuisines"].fillna("Not Available", inplace=True)

[ ]:
```

Check duplicate rows:

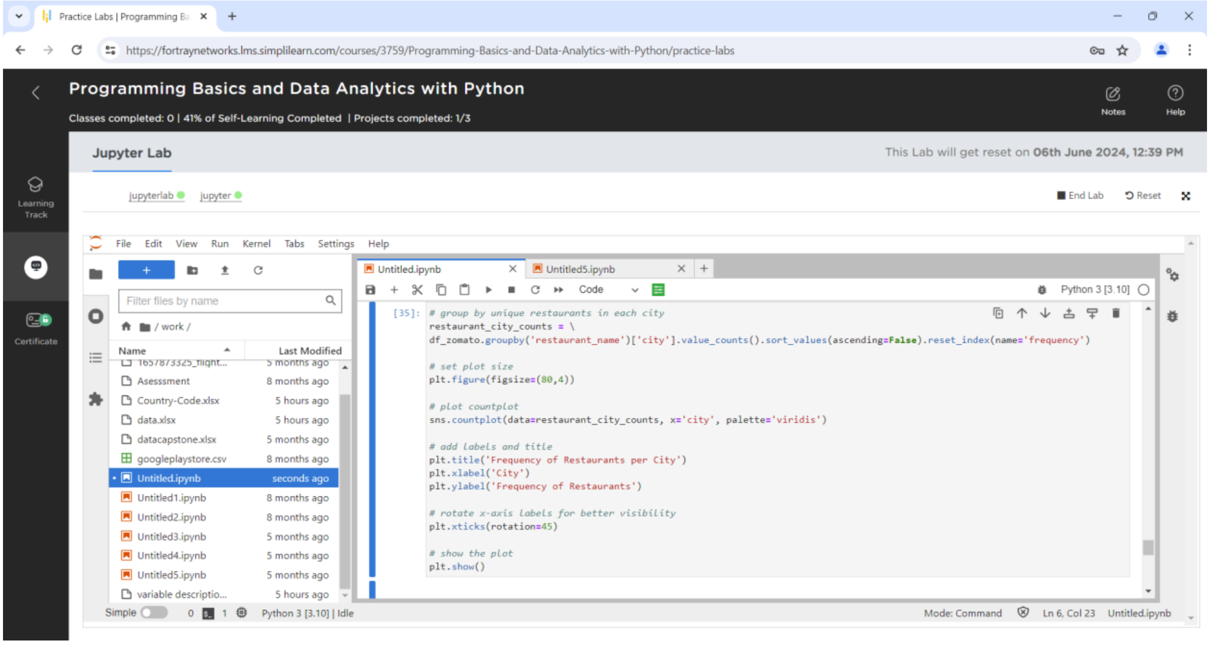


```
[27]: # duplicate rows
      df_zomato[df_zomato.duplicated()].head()

[27]:  restaurant_id  restaurant_name  country_code  city  address  locality  locality_verbose  longitude  latitude  cuisines  average_cost_for_tw

[28]: df_zomato = df_zomato.drop_duplicates()

[ ]:
```

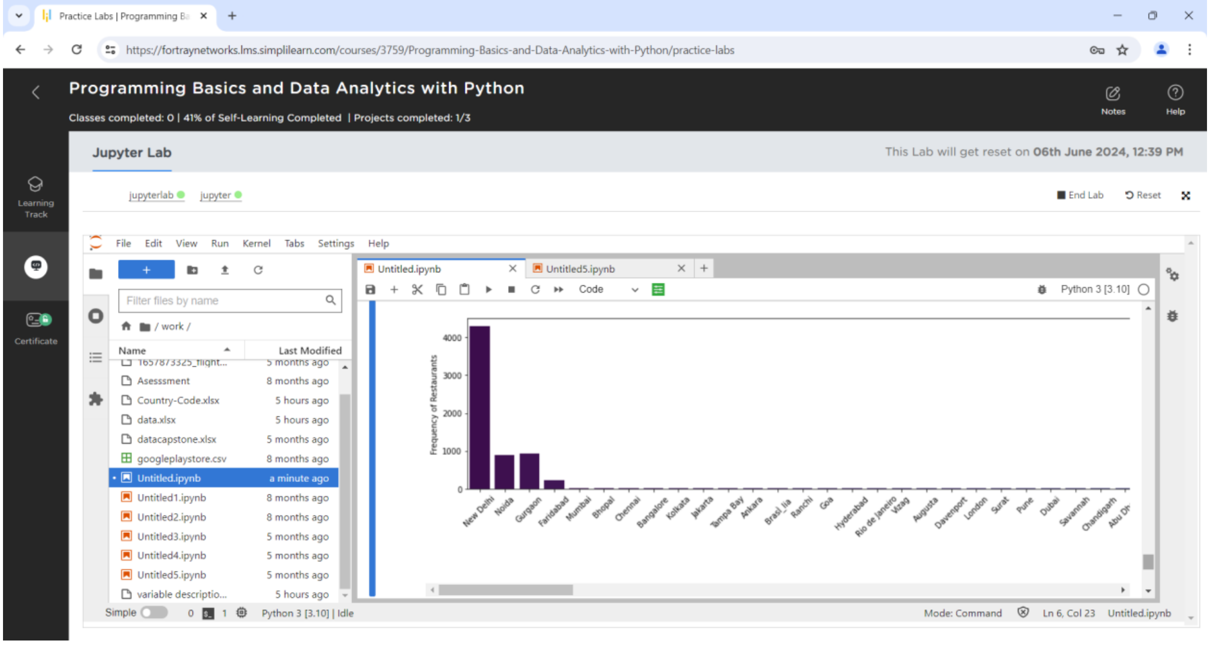
Explore the geographical distribution of the restaurants, finding out the cities with maximum / minimum number of restaurants.



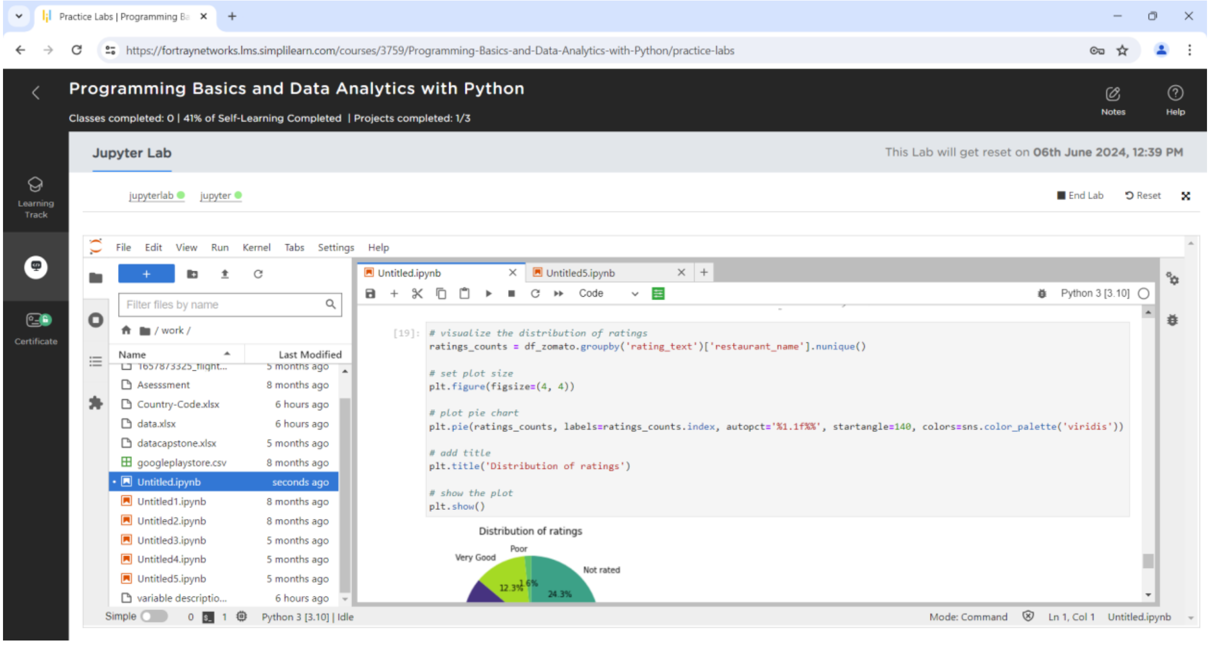
```
[29]: # check value counts
      df_zomato["city"].value_counts()

[29]: New Delhi            5473
      Gurgaon              1118
      Noida                1080
      Faridabad             251
      Ghaziabad              25
      Lucknow                21
      Bhubaneshwar           21
      Ahmedabad              21
      Amritsar               21
      Guwahati               21
      Allahabad              20
      Macon                  20
      Rest of Hawaii         20
      Savannah               20
      Sioux City             20
      Sharjah                20
      Tampa Bay              20
      Cape Town              20
```


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Name
Last Modified
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx 5 hours ago
data.xlsx 5 hours ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb seconds ago
Untitled1.ipynb 8 months ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled4.ipynb 5 months ago
Untitled5.ipynb 5 months ago
variable descriptio... 5 hours ago
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Simple 0 1 Python 3 [3.10] | Idle
Mode: Command
Ln 6, Col 23
Untitled.ipynb

```
[35]: # group by unique restaurants in each city
      restaurant_city_counts = \
      df_zomato.groupby('restaurant_name')['city'].value_counts().sort_values(ascending=False).reset_index(name='frequency')

      # set plot size
      plt.figure(figsize=(80,4))

      # plot countplot
      sns.countplot(data=restaurant_city_counts, x='city', palette='viridis')

      # add labels and title
      plt.title('Frequency of Restaurants per City')
      plt.xlabel('City')
      plt.ylabel('Frequency of Restaurants')

      # rotate x-axis labels for better visibility
      plt.xticks(rotation=45)

      # show the plot
      plt.show()
```


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Name
Last Modified
1657873325_flight... 5 months ago
Asesssment 8 months ago
Country-Code.xlsx 5 hours ago
data.xlsx 5 hours ago
datacapstone.xlsx 5 months ago
googleplaystore.csv 8 months ago
Untitled.ipynb a minute ago
Untitled1.ipynb 8 months ago
Untitled2.ipynb 8 months ago
Untitled3.ipynb 5 months ago
Untitled4.ipynb 5 months ago
Untitled5.ipynb 5 months ago
variable descriptio... 5 hours ago
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Frequency of Restaurants
4000
3000
2000
1000
0
New Delhi
Noida
Gurgaon
Faridabad
Mumbai
Bhopal
Chennai
Bangalore
Kolkata
Jakarta
Tampa Bay
Ankara
Brasí_lia
Ranchi
Goa
Hyderabad
Rio de Janeiro
Vizag
Augusta
Davenport
London
Surat
Pune
Dubai
Savannah
Chandigarh
Abu Dh
Simple 0 1 Python 3 [3.10] | Idle
Mode: Command
Ln 6, Col 23
Untitled.ipynb

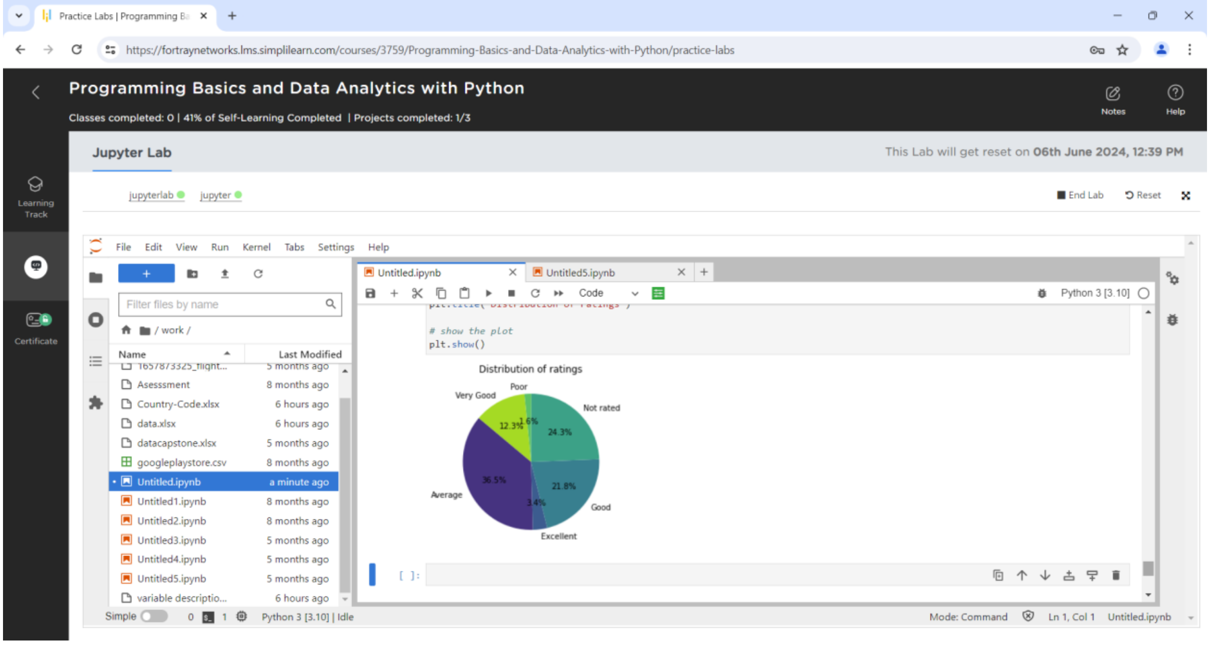
Explore how ratings are distributed overall.


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Learning Track
Certificate
jupyterlab
jupyter
End Lab
Reset
File Edit View Run Kernel Tabs Settings Help
Filter files by name
/ work /
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled.ipynb

Name	Last Modified
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	6 hours ago
data.xlsx	6 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago
Untitled5.ipynb	5 months ago
variable descriptio...	6 hours ago

```
[19]: # visualize the distribution of ratings
      ratings_counts = df_zomato.groupby('rating_text')['restaurant_name'].nunique()

      # set plot size
      plt.figure(figsize=(4, 4))

      # plot pie chart
      plt.pie(ratings_counts, labels=ratings_counts.index, autopct='%1.1f%%', startangle=140, colors=sns.color_palette('viridis'))

      # add title
      plt.title('Distribution of ratings')

      # show the plot
      plt.show()
```

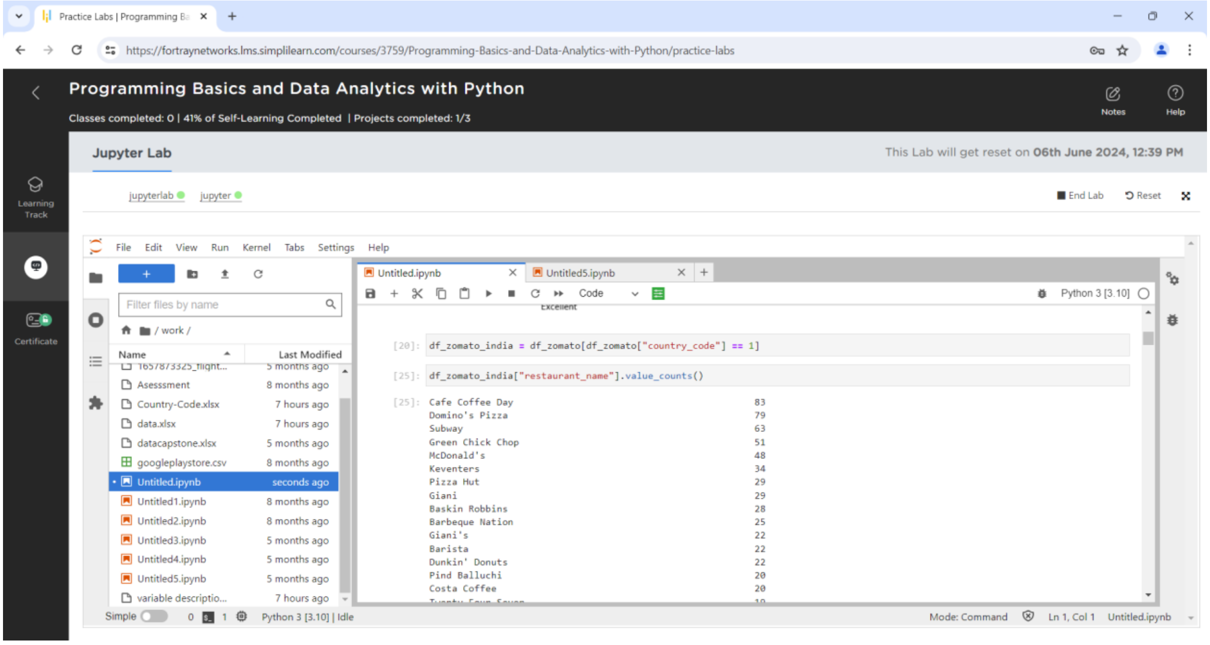
Distribution of ratings
Poor
Very Good
Not rated
12.3%
1.6%
24.3%


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
a minute ago
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1

```
      # show the plot
      plt.show()
```

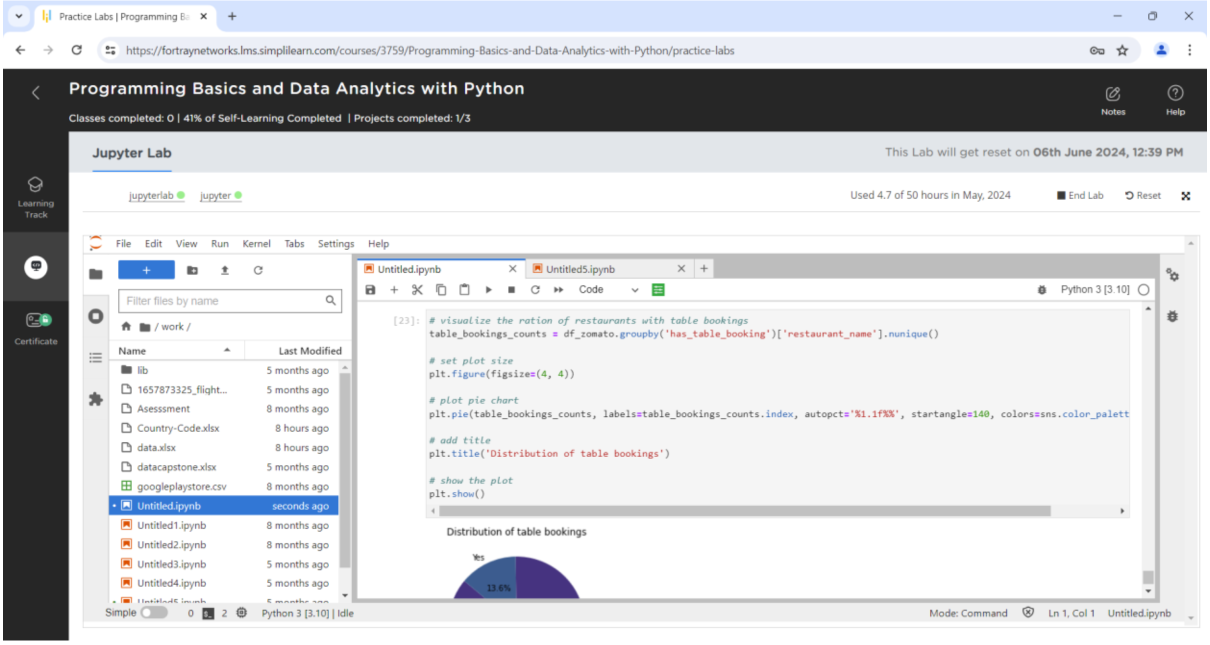
Distribution of ratings
Poor
Very Good
Not rated
12.3%
1.6%
24.3%
36.5%
21.8%
3.4%
Average
Good
Excellent
[ ]:

Explore the franchise with most national presence:

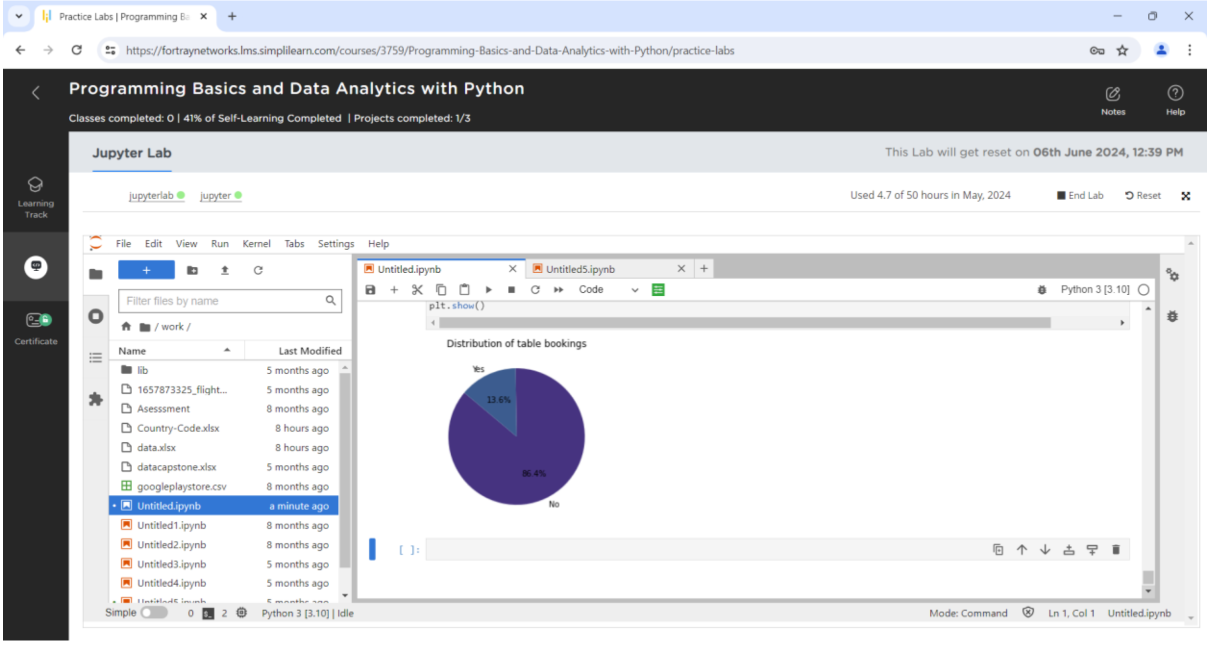

Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Untitled.ipynb
Untitled5.ipynb
Python 3 [3.10]

```
[20]: df_zomato_india = df_zomato[df_zomato["country_code"] == 1]

[25]: df_zomato_india["restaurant_name"].value_counts()

[25]: Cafe Coffee Day                                                 83
      Domino's Pizza                                                  79
      Subway                                                          63
      Green Chick Chop                                                51
      McDonald's                                                      48
      Keventers                                                       34
      Pizza Hut                                                       29
      Giani                                                           29
      Baskin Robbins                                                  28
      Barbeque Nation                                                 25
      Giani's                                                         22
      Barista                                                         22
      Dunkin' Donuts                                                  22
      Pind Balluchi                                                   20
      Costa Coffee                                                    20
```

What is the ratio between restaurants that allow table booking vs that do not allow table booking?


Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
Used 4.7 of 50 hours in May, 2024
Untitled.ipynb
Untitled5.ipynb
Python 3 [3.10]

```
[23]: # visualize the ration of restaurants with table bookings
      table_bookings_counts = df_zomato.groupby('has_table_booking')['restaurant_name'].nunique()

      # set plot size
      plt.figure(figsize=(4, 4))

      # plot pie chart
      plt.pie(table_bookings_counts, labels=table_bookings_counts.index, autopct='%1.1f%%', startangle=140, colors=sns.color_palett

      # add title
      plt.title('Distribution of table bookings')

      # show the plot
      plt.show()
```



https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs

Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

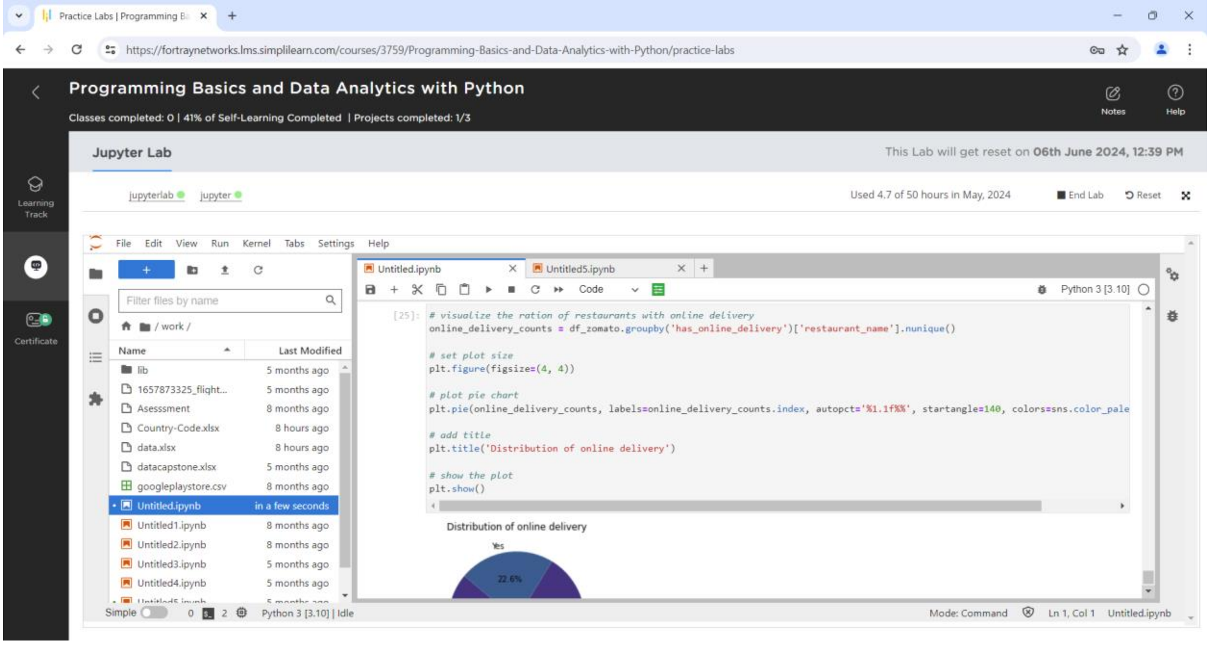
This Lab will get reset on 06th June 2024, 12:39 PM

Used 4.7 of 50 hours in May, 2024

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	8 hours ago
data.xlsx	8 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	a minute ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
plt.show()
```

What is the percentage of restaurants providing online delivery?



https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs

Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

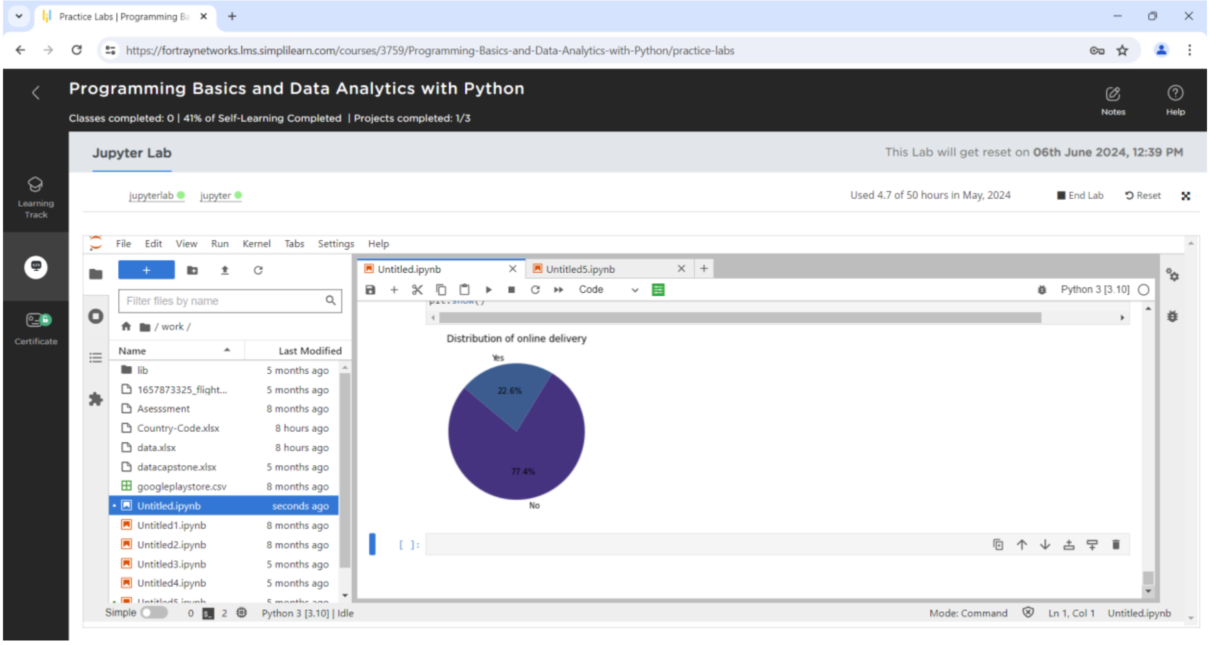
Jupyter Lab

This Lab will get reset on 06th June 2024, 12:39 PM

Used 4.7 of 50 hours in May, 2024

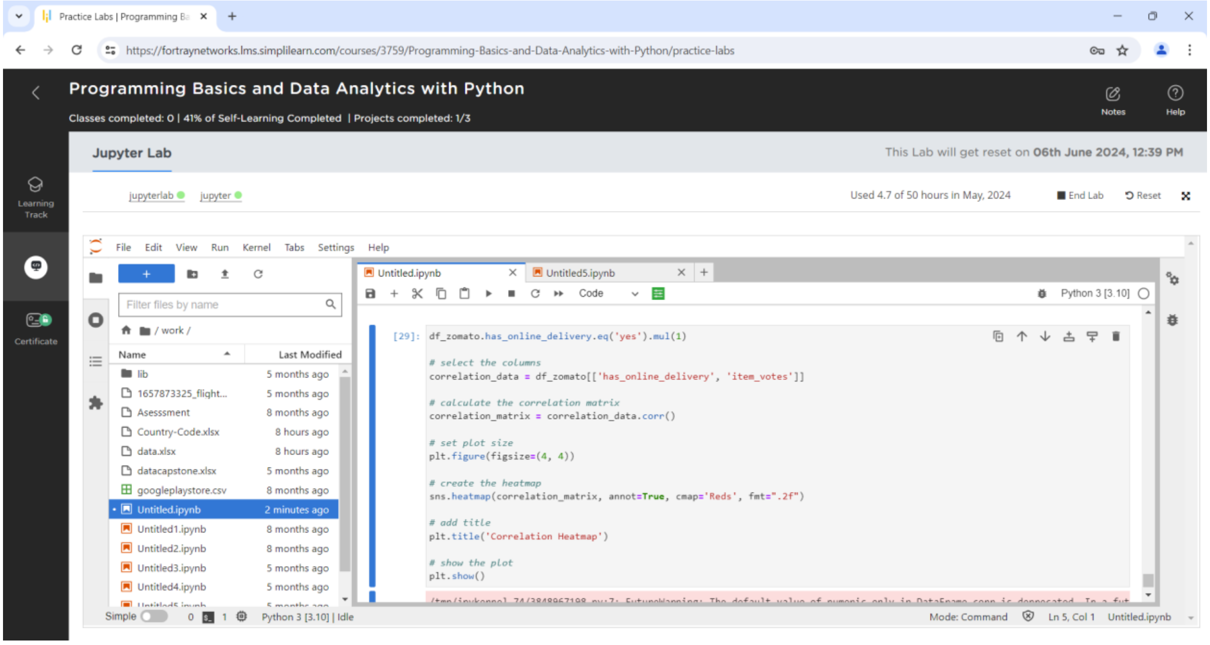
Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	8 hours ago
data.xlsx	8 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	in a few seconds
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
[25]: # visualize the ration of restaurants with online delivery
online_delivery_counts = df_zomato.groupby('has_online_delivery')['restaurant_name'].nunique()

# set plot size
plt.figure(figsize=(4, 4))

# plot pie chart
plt.pie(online_delivery_counts, labels=online_delivery_counts.index, autopct='%1.1f%%', startangle=140, colors=sns.color_pale

# add title
plt.title('Distribution of online delivery')

# show the plot
plt.show()
```


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
Used 4.7 of 50 hours in May, 2024
End Lab
Reset
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Filter files by name
/ work /
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Distribution of online delivery
Yes
22.6%
77.4%
No
[ ]:
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 1, Col 1
Untitled.ipynb

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	8 hours ago
data.xlsx	8 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

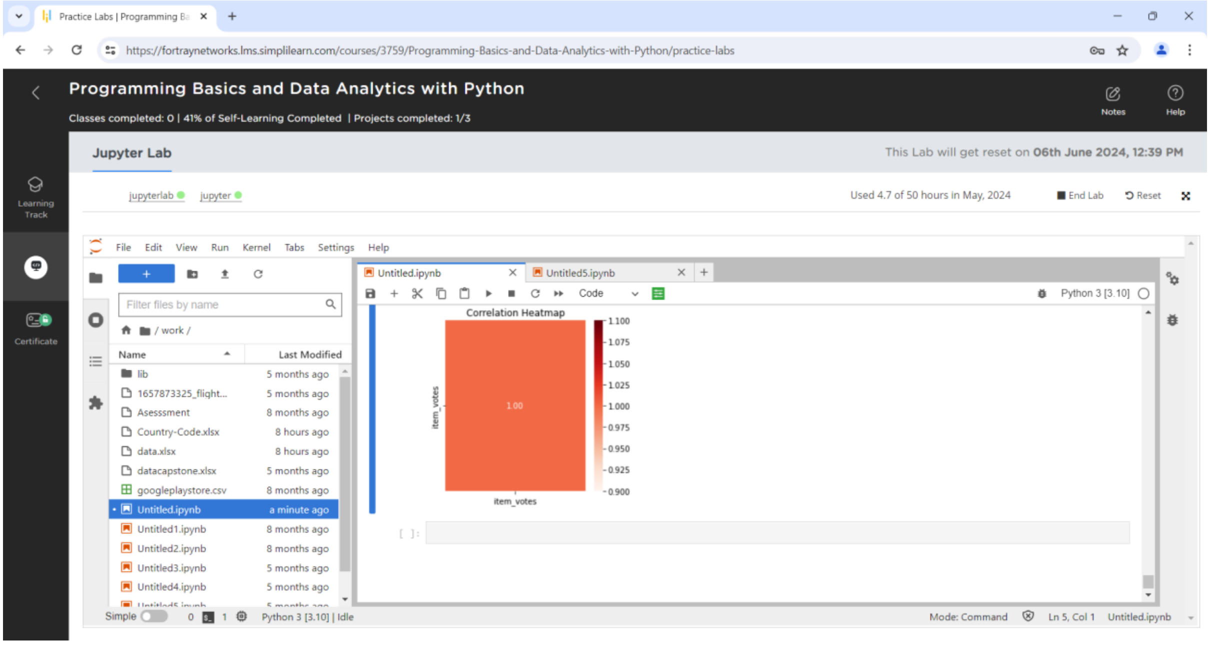
Is there a difference in no. of votes for the restaurants that deliver and the restaurant that don’t?


Practice Labs | Programming Ba
https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs
Programming Basics and Data Analytics with Python
Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3
Notes
Help
Learning Track
Certificate
Jupyter Lab
This Lab will get reset on 06th June 2024, 12:39 PM
jupyterlab
jupyter
Used 4.7 of 50 hours in May, 2024
End Lab
Reset
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Filter files by name
/ work /
Untitled.ipynb
Untitled5.ipynb
Code
Python 3 [3.10]
Simple
Python 3 [3.10] | Idle
Mode: Command
Ln 5, Col 1
Untitled.ipynb

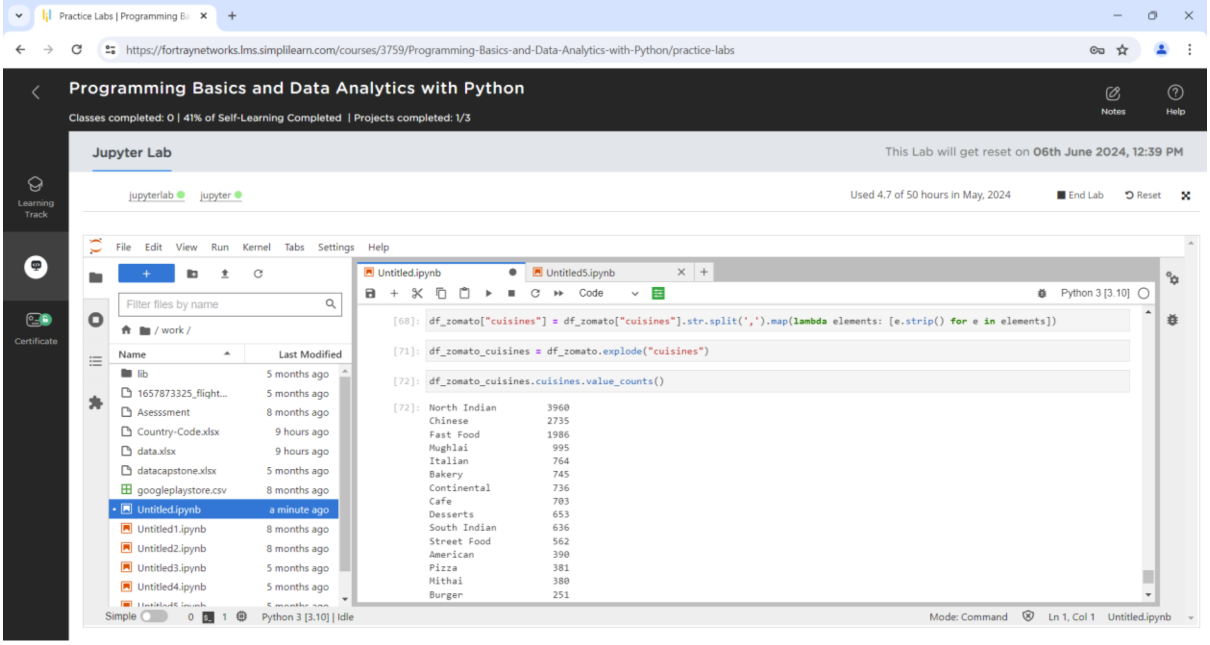
Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	8 hours ago
data.xlsx	8 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	2 minutes ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
[29]: df_zomato.has_online_delivery.eq('yes').mul(1)

      # select the columns
      correlation_data = df_zomato[['has_online_delivery', 'item_votes']]

      # calculate the correlation matrix
      correlation_matrix = correlation_data.corr()

      # set plot size
      plt.figure(figsize=(4, 4))

      # create the heatmap
      sns.heatmap(correlation_matrix, annot=True, cmap='Reds', fmt=".2f")

      # add title
      plt.title('Correlation Heatmap')

      # show the plot
      plt.show()
```

```
/tmp/ipykernel_74/3848967198.py:7: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a fut
```

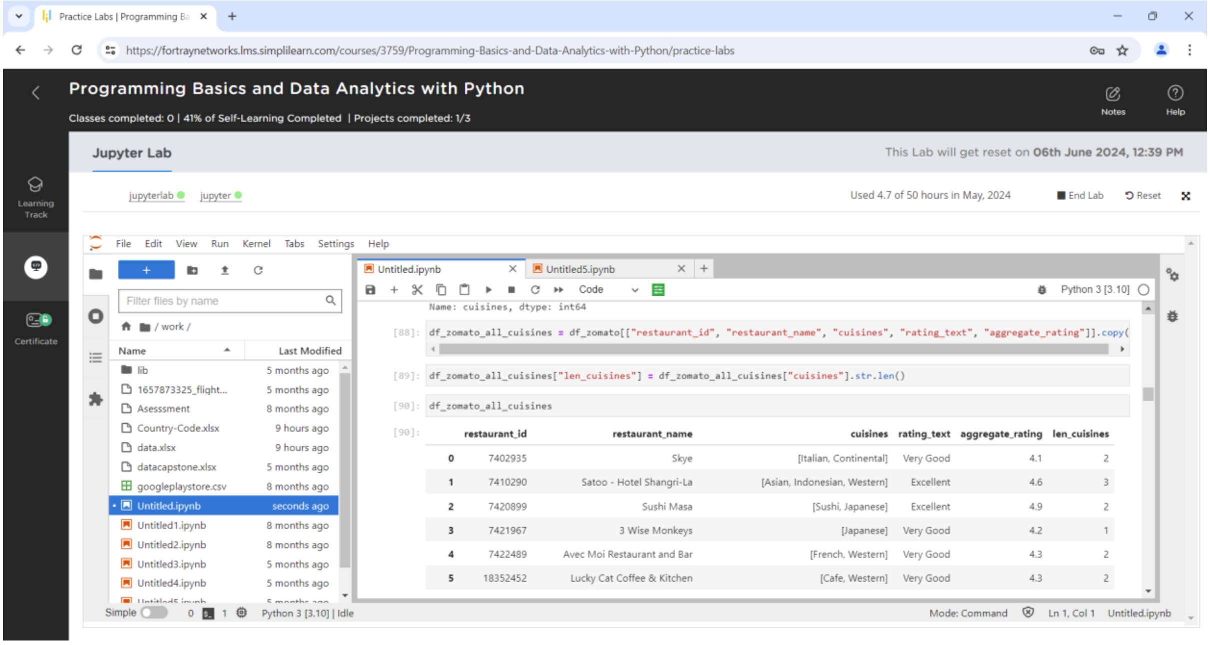


What are the top 10 cuisines served across cities?



```
[68]: df_zomato["cuisines"] = df_zomato["cuisines"].str.split(',').map(lambda elements: [e.strip() for e in elements])

[71]: df_zomato_cuisines = df_zomato.explode("cuisines")

[72]: df_zomato_cuisines.cuisines.value_counts()

[72]: North Indian        3960
      Chinese             2735
      Fast Food           1986
      Mughlai              995
      Italian              764
      Bakery               745
      Continental          736
      Cafe                 703
      Desserts             653
      South Indian         636
      Street Food          562
      American             390
      Pizza                381
      Mithai               380
      Burger               251
```

What is the maximum and minimum no. of cuisines that a restaurant serves? Also, what is the relationship between No. of cuisines served and Ratings



Practice Labs | Programming Ba

https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs

Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

This Lab will get reset on **06th June 2024, 12:39 PM**

Used 4.7 of 50 hours in May, 2024

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	9 hours ago
data.xlsx	9 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
Name: cuisines, dtype: int64
```

```
[88]: df_zomato_all_cuisines = df_zomato[["restaurant_id", "restaurant_name", "cuisines", "rating_text", "aggregate_rating"]].copy(
```

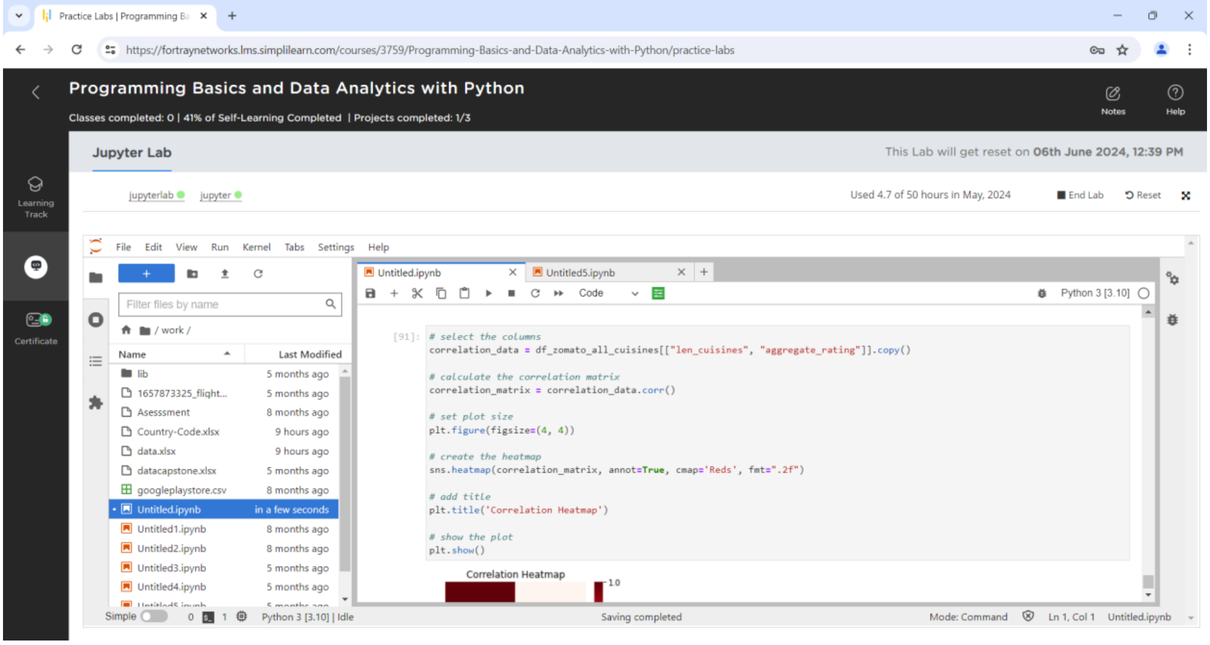
```
[89]: df_zomato_all_cuisines["len_cuisines"] = df_zomato_all_cuisines["cuisines"].str.len()
```

```
[90]: df_zomato_all_cuisines
```

[90]:

	restaurant_id	restaurant_name	cuisines	rating_text	aggregate_rating	len_cuisines
0	7402935	Skye	[Italian, Continental]	Very Good	4.1	2
1	7410290	Satoo - Hotel Shangri-La	[Asian, Indonesian, Western]	Excellent	4.6	3
2	7420899	Sushi Masa	[Sushi, Japanese]	Excellent	4.9	2
3	7421967	3 Wise Monkeys	[Japanese]	Very Good	4.2	1
4	7422489	Avec Moi Restaurant and Bar	[French, Western]	Very Good	4.3	2
5	18352452	Lucky Cat Coffee & Kitchen	[Cafe, Western]	Very Good	4.3	2

0 1 Python 3 [3.10] | Idle Mode: Command Ln 1, Col 1 Untitled.ipynb



Practice Labs | Programming Ba

https://fortraynetworks.lms.simplilearn.com/courses/3759/Programming-Basics-and-Data-Analytics-with-Python/practice-labs

Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

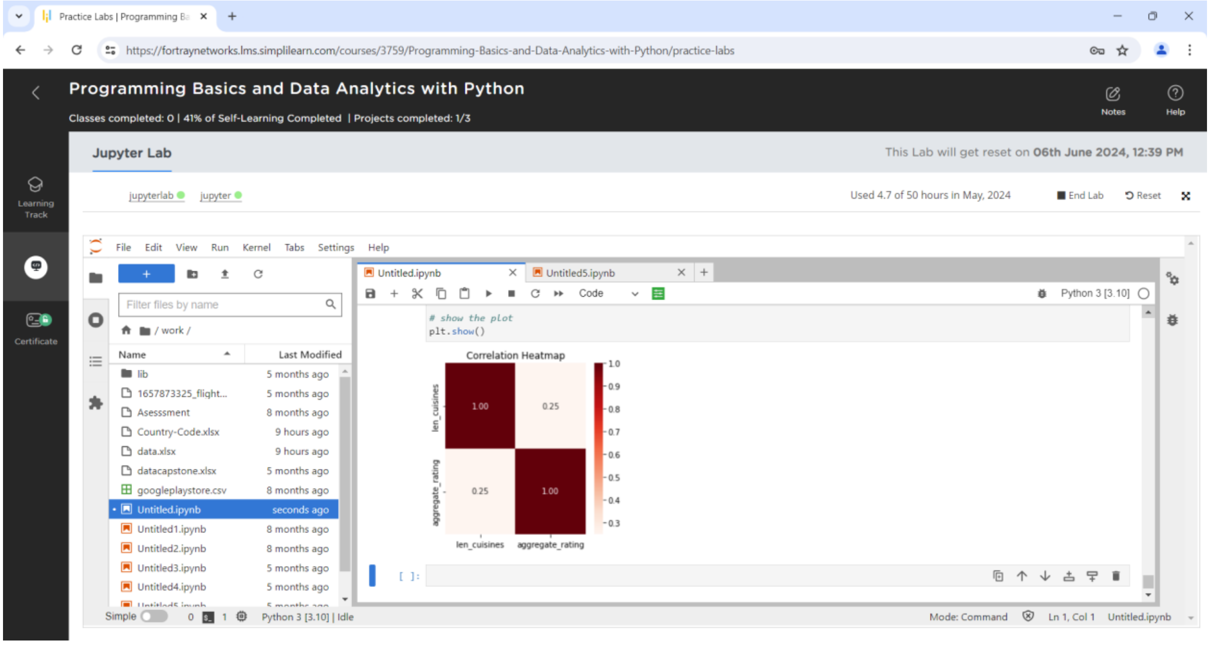
This Lab will get reset on 06th June 2024, 12:39 PM

Used 4.7 of 50 hours in May, 2024

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asessment	8 months ago
Country-Code.xlsx	9 hours ago
data.xlsx	9 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	in a few seconds
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
[91]: # select the columns
correlation_data = df_zomato_all_cuisines[["len_cuisines", "aggregate_rating"]].copy()

# calculate the correlation matrix
correlation_matrix = correlation_data.corr()

# set plot size
plt.figure(figsize=(4, 4))

# create the heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='Reds', fmt=".2f")

# add title
plt.title('Correlation Heatmap')

# show the plot
plt.show()
```

0 1 Python 3 [3.10] | Idle Saving completed Mode: Command Ln 1, Col 1 Untitled.ipynb



Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

This Lab will get reset on 06th June 2024, 12:39 PM

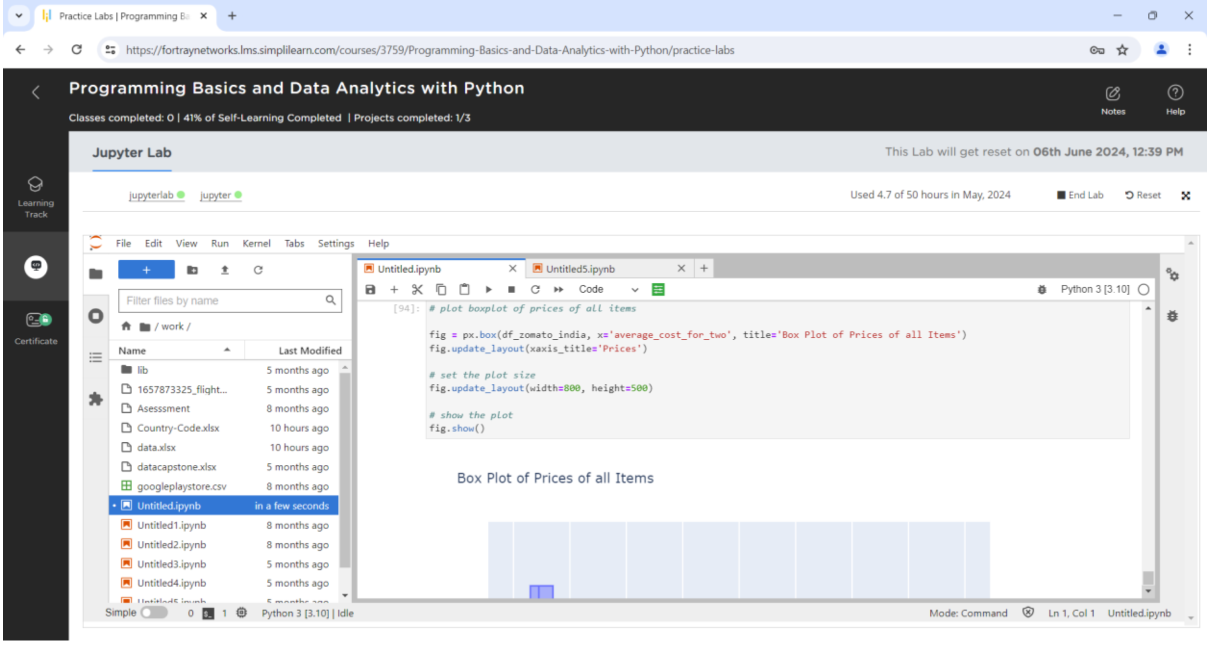
Used 4.7 of 50 hours in May, 2024

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	9 hours ago
data.xlsx	9 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	seconds ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
# show the plot
plt.show()
```

```
[ ]:
```

Discuss the cost vs the other variables.



Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

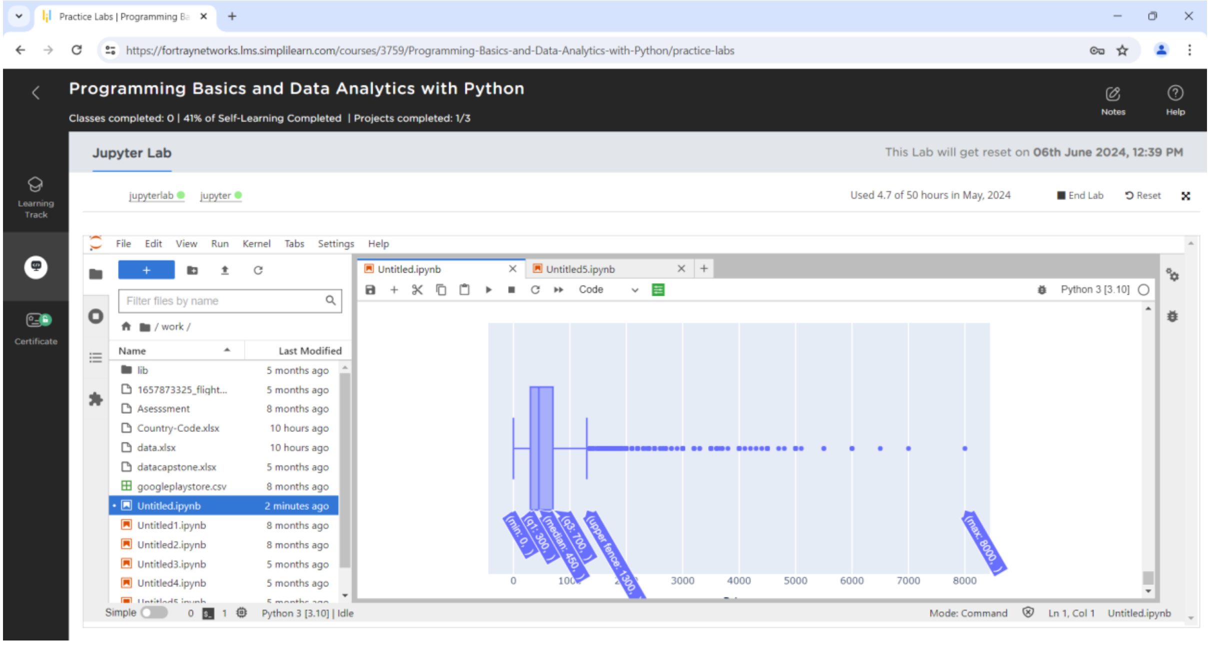
Jupyter Lab

This Lab will get reset on 06th June 2024, 12:39 PM

Used 4.7 of 50 hours in May, 2024

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	10 hours ago
data.xlsx	10 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	in a few seconds
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

```
[94]: # plot boxplot of prices of all items

      fig = px.box(df_zomato_india, x='average_cost_for_two', title='Box Plot of Prices of all Items')
      fig.update_layout(xaxis_title='Prices')

      # set the plot size
      fig.update_layout(width=800, height=500)

      # show the plot
      fig.show()
```



Programming Basics and Data Analytics with Python

Classes completed: 0 | 41% of Self-Learning Completed | Projects completed: 1/3

Jupyter Lab

This Lab will get reset on 06th June 2024, 12:39 PM

Used 4.7 of 50 hours in May, 2024

/ work /

Name	Last Modified
lib	5 months ago
1657873325_flight...	5 months ago
Asesssment	8 months ago
Country-Code.xlsx	10 hours ago
data.xlsx	10 hours ago
datacapstone.xlsx	5 months ago
googleplaystore.csv	8 months ago
Untitled.ipynb	2 minutes ago
Untitled1.ipynb	8 months ago
Untitled2.ipynb	8 months ago
Untitled3.ipynb	5 months ago
Untitled4.ipynb	5 months ago

It revealed that the median price for all items in India is 450 rupees, with the majority falling within the range of 300 to 700 rupees. However, several outliers are present, indicating dishes with significantly higher prices, with one reaching as high as 8000 rupees.

Explain the factors in the data that may have an effect on ratings e.g. No. of cuisines, cost, delivery option etc.

Dining Type Analysis offered valuable insights into the customer's inclination towards dining or delivery services. Versatility in service, offering both dining and delivery options, proved to be a key factor in maximizing engagement.

The City Analysis revealed the best-suited locations for opening a restaurant, considering factors like engagement and competition. New Delhi surfaced as the busiest location, having maximum number of competitors.