Data Science – Data Visualisation with Altair

• Go to: https://www.kaggle.com/datasets
• Select a dataset
• Then create a data dashboard using Altair or Tableau https://public.tableau.com/en-us/s/
• Then create a markdown cell explain why you have decided on the design choices, what influenced your decisions and what insights have you found from the data.

```
import altair as alt
import pandas as pd
import numpy as np

dataset = pd.read_csv (r'C:\Users\sudip\holiday_destination.csv')
print (dataset)
```

```
   holiday_destination most_visited_city      country  all_inclusive_package  \
0          Burj Khalifa             Dubai          UAE                      20
1             MGM Grand         Las Vegas          USA                      15
2          Tower Bridge            London           UK                       9
3         Tokyo Skytree             Tokyo        Japan                      16
4         Palolem Beach               Goa        India                      21
5             Khao Rang            Phuket     Thailand                      12
6           Mount Batur              Bali    Indonesia                       7
7    Escadaria Selarón    Rio de Janeiro       Brazil                      25
8       Grouse Mountain         Vancouver       Canada                      18
9            Lovrijenac         Dubrovnik      Croatia                       8
10            Red Beach         Santorini       Greece                      19
11    Düden Waterfalls           Antalya       Turkey                      15
12            Colosseum              Rome        Italy                      17
13         Opera House            Sydney    Australia                      10
14           Vondelpark         Amsterdam  Netherlands                       6

   feedback_score  hotel_star_rating  average_review_score
0       Excellent                4.7                   9.6
1       Very Good                4.0                   8.9
2       Excellent                4.7                   8.6
3       Very Good                4.4                   8.4
4       Excellent                4.6                   8.2
5       Very Good                4.5                   7.3
6       Excellent                4.6                   8.2
7       Excellent                4.6                   8.2
8       Excellent                4.5                   7.5
9       Excellent                4.7                   8.6
10      Very Good                4.1                   6.9
11      Excellent                4.7                   8.3
12      Excellent                4.7                   7.5
13      Excellent                4.7                   9.3
14      Excellent                4.7                   8.3
```

```
# Size of Dataset
print(dataset.shape)
```

(15, 7)

# Top 5 Rows
dataset.head(5)

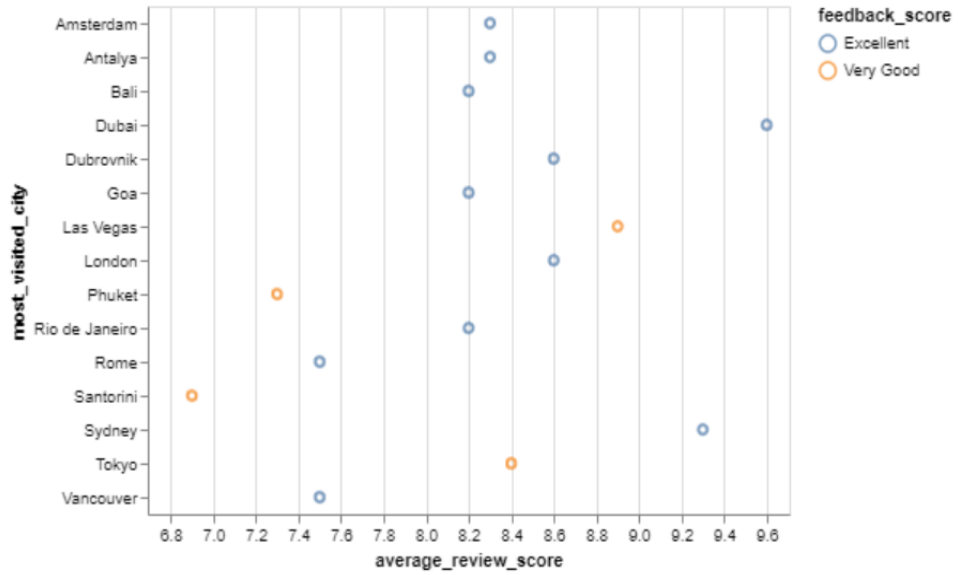
	holiday_destination	most_visited_city	country	all_inclusive_package	feedback_score	hotel_star_rating	average_review_score
0	Burj Khalifa	Dubai	UAE	20	Excellent	4.7	9.6
1	MGM Grand	Las Vegas	USA	15	Very Good	4.0	8.9
2	Tower Bridge	London	UK	9	Excellent	4.7	8.6
3	Tokyo Skytree	Tokyo	Japan	16	Very Good	4.4	8.4
4	Palolem Beach	Goa	India	21	Excellent	4.6	8.2

# The statistical details of the dataset using describe().
dataset.describe()

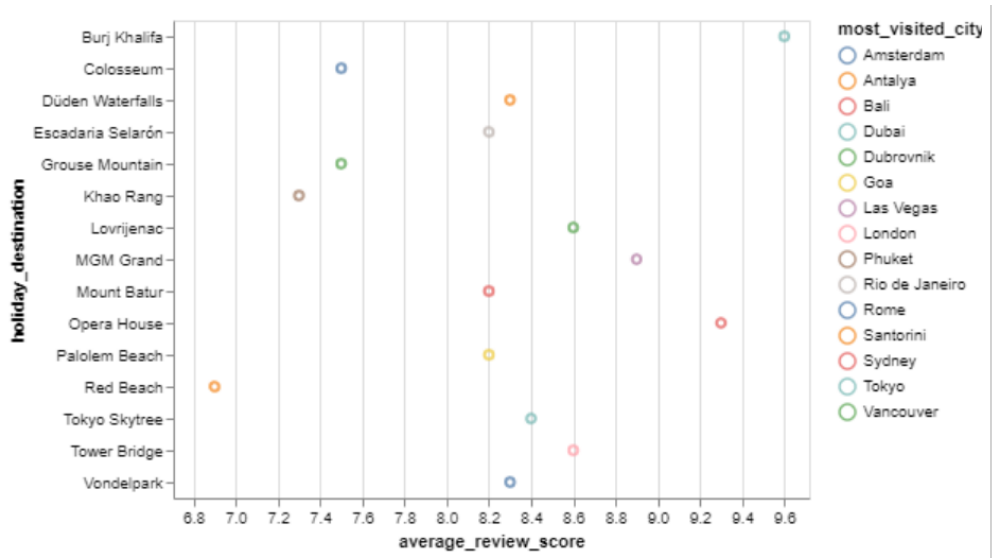
	all_inclusive_package	hotel_star_rating	average_review_score
count	15.000000	15.000000	15.000000
mean	14.533333	4.546667	8.253333
std	5.680376	0.223180	0.730818
min	6.000000	4.000000	6.900000
25%	9.500000	4.500000	7.850000
50%	15.000000	4.600000	8.300000
75%	18.500000	4.700000	8.600000
max	25.000000	4.700000	9.600000

Plotting and Creating Charts

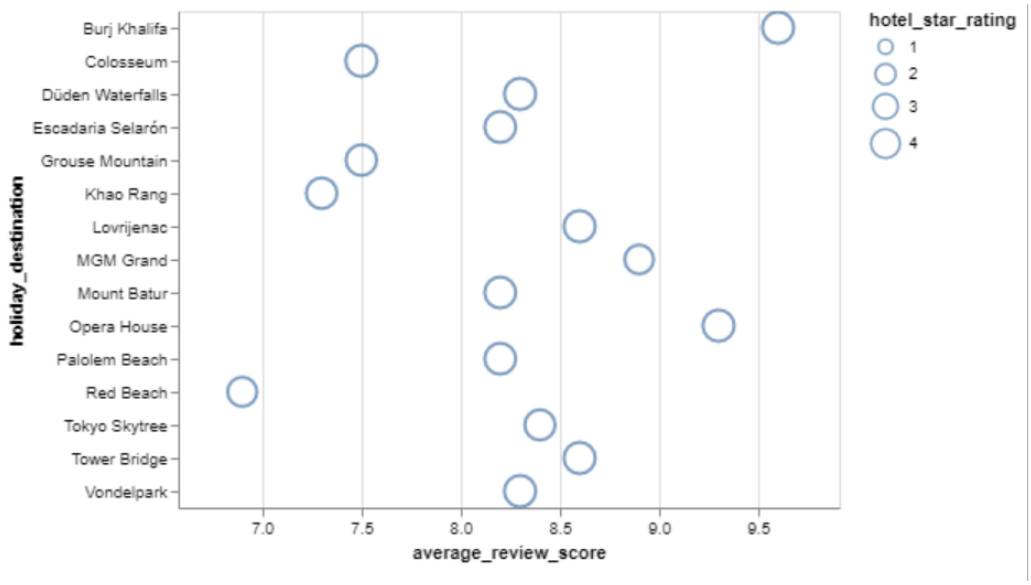
```
alt.Chart(dataset).mark_point().encode(
    x='average_review_score',
    y='most_visited_city',
    color='feedback_score'
).interactive()
```



```
alt.Chart(dataset).mark_point().encode(
    x='average_review_score',
    y='holiday_destination',
    color='most_visited_city'
).interactive()
```

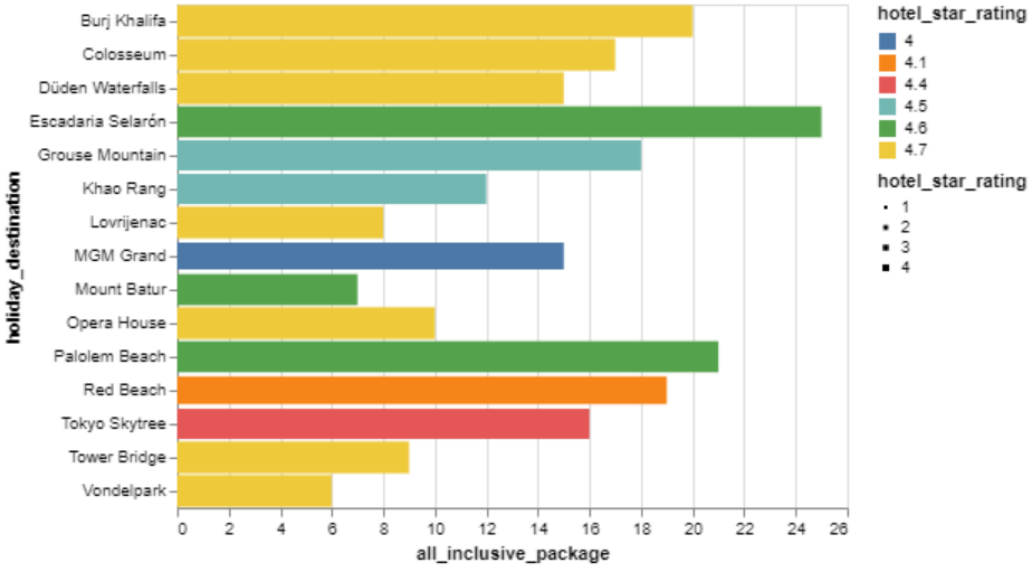



```
alt.Chart(dataset).mark_point().encode(
    x='average_review_score',
    y='holiday_destination',
    size='hotel_star_rating',
).interactive()
```

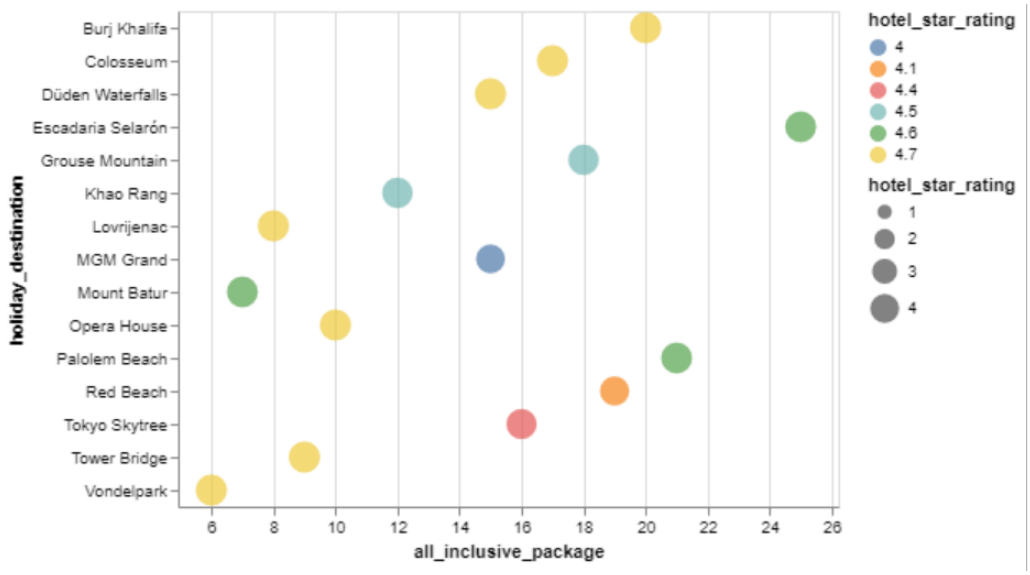



```
barPlot = alt.Chart(dataset).mark_bar().encode(
    x='all_inclusive_package',
    y='holiday_destination',
    size ='hotel_star_rating',
    color='hotel_star_rating:N'
).interactive()
```

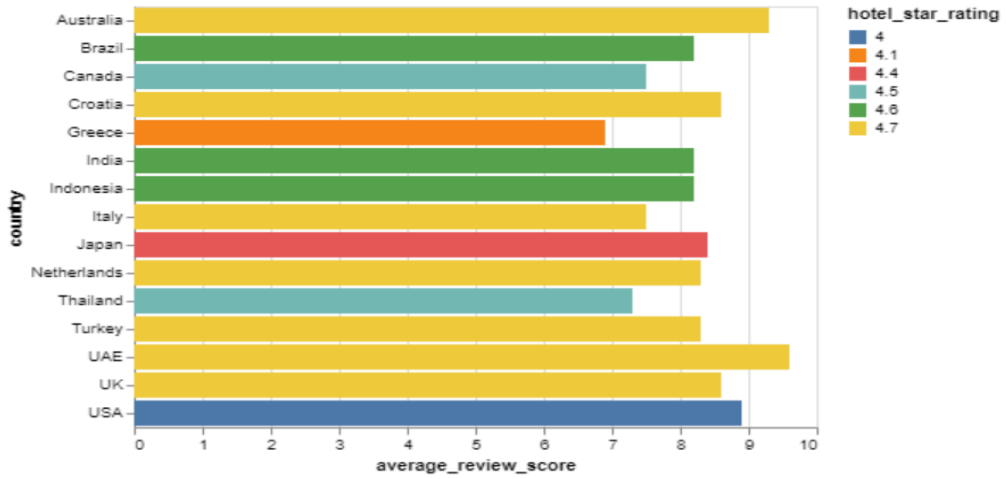
barplot




```
circlePlot = alt.Chart(dataset).mark_circle().encode(
    x='all_inclusive_package',
    y='holiday_destination',
    size = 'hotel_star_rating',
    color='hotel_star_rating:N',
    tooltip=['country','feedback_score'],
).interactive()
circlePlot
```




```
stackbarPlot = alt.Chart(dataset).mark_bar().encode(
    x='average_review_score',
    y='country',
    color='hotel_star_rating:N',
    tooltip=['holiday_destination', 'hotel_star_rating']
)
stackbarPlot
```



I found Data Visualisation module very challenging, initially found various datasets on kaggle and tried to use Tableau, however even after using the guide and YouTube tips, I still found it quite difficult to use it. Then, I attempted the Altair method, so far it's much easier to understand. I chose an easier dataset, following the holiday destination homework in Pandas but adding more details then converting it into a CSV file. When I tried importing my file, I kept receiving a unicode error, after some research and going through my files, I've found that converting my regular CSV file to a CSV-UTF-8 fixed the unicode error I kept receiving on Jupyter Notebook. After researching on some graphs, I really loved the look of circlePlot and stackbarPlot, however I knew because I don't have much data to work with, for the most parts it will look incomplete. If I were to do this again, I would keep the feedback_score and hotel_star_rating balanced rather than just keeping the values really high, so the graphs would not feel incomplete.