

**Assignment 1: Visualisation**

*University of Hertfordshire*

*Department of Physics, Astronomy and Mathematics*

*School of Physics, Engineering and Computer Science*

***Module:*** 7PAM2000-0105-2022 - Applied Data Science 1

***Submitted To:*** Ralf Napiwotzki (Module leader)

***Submitted By:*** Vasthav Dandumenu (22032815)

***Date of Submission:*** March 5th, 2023

Icon

Description automatically generated with medium confidence

***(1) Line plot visualization:***

***State–Wise Covid Analysis of India, March 2020.***

* ***Link to Git Repository:*** [Vasthav\_Dandumenu\_22032815](https://github.com/vasthavd/Vasthav_Dandumenu_22032815)
* ***Link to code file:*** [Question\_1\_Covid\_State\_Analysis\_Line\_Plot.py](https://github.com/vasthavd/Vasthav_Dandumenu_22032815/blob/ae5b6c6db50e68b5c3d4849ea2fb24a92ecb4c20/Question_1_Covid_State_Analysis_Line_Plot.py)
* ***Link to dataset:*** [March.xlsx](https://docs.google.com/spreadsheets/d/1R9opFJAFmCWGj4tYyEOHBYmLYn9cWj8V/edit#gid=1790449400) or [MarchGit.xlsx](https://github.com/vasthavd/Vasthav_Dandumenu_22032815/blob/ae5b6c6db50e68b5c3d4849ea2fb24a92ecb4c20/March.xlsx)
* ***Choice of plot:*** I have chosen the line plot to visualize this Covid data, as the data is based on covid cases from time to time. As in a line plot, we can plot specific points connected to each other which in turn will be helpful to analyze the pattern and trend of overall growth or decline in the cases over a time period comparatively better than on a scatter plot.
* ***Code text file and plots:*** Attaching code text file & March dataset.

 

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

* ***Description of plots:*** The above line plots have been plotted for the March Covid data of 2020 for the states of Maharashtra and Kerala from India. With

x-axis showing the dates of the month of March and y-axis displaying the number of covid cases reported for each type. As shown in the legend Confirmed Indian National cases are shown by blue line. Similarly, Confirmed Foreign National, Cured and Deaths have been represented by orange, green and red coloured lines respectively. The above plots have been plotted from modified data for a single month. The original dataset contains information for three months from January to March 2020.

* ***Insights:*** From the above plots we can derive the following insights:
  1. In both states Indians were contracting Covid at a higher rate than foreigners, and cases were constantly rising throughout the month.
  2. Kerala recorded no deaths, whereas Maharashtra recorded one from the 16th to the 21st of March 2020.
  3. In both states the recovery rate was low although Kerala’s recovery was comparatively faster than Maharashtra’s.
  4. Maharashtra didn’t record any Covid cases during the first week of March.

***(2) Pie plot visualization:***

***Summer Olympics Medal Analysis, 1896 - 2008.***

* ***Link to Git Repository:*** [Vasthav\_Dandumenu\_22032815](https://github.com/vasthavd/Vasthav_Dandumenu_22032815)
* ***Link to code file:*** [Question\_2\_Summer\_Olympics\_Pie\_Plot.py](https://github.com/vasthavd/Vasthav_Dandumenu_22032815/blob/ae5b6c6db50e68b5c3d4849ea2fb24a92ecb4c20/Question_2_Summer_Olympics_Pie_Plot.py)
* ***Link to dataset:*** [Summer\_Olympic\_medallists\_1896-2008.xlsx](https://public.tableau.com/app/sample-data/Summer_Olympic_medallists_1896-2008.xlsx)
* ***Choice of plot:*** I have chosen the pie plot to visualize the summer Olympics data, as I wanted to show the percentages of women’s and men’s medals in different sports altogether for all the participating countries and as the pie plot is circular in style, this would be the best way to graphically represents the percentage of categories.
* ***Code text file and plots:*** Attaching code text file.



Chart, pie chart

Description automatically generatedChart, pie chart

Description automatically generated

* ***Description of plots:*** The above pie plots have been plotted for the Summer Olympics data of 1896 – 2008, taking into consideration the performance of different gender people receiving medals in a particular sport during the whole event. The above plots are a result of inputs to plot medal analysis of Men’s and Women’s in the sport of football. Different types of medals have been shown on the plot along with the percentages of the specific medals received by a particular gender in a single sport for all the countries. And this can be plotted for any desired gender or sport from the data.
* ***Insights:*** From the above plots we can derive the following insights:
  1. In both plots we can notice men and women received a nearly similar number of medals in all three categories namely gold, silver, and bronze.
  2. Men received more gold medals than women and women received more bronze and silver medals than men in football.
  3. Both men and women received more bronze medals in football.
  4. Men and women performed very similarly to each other on a sport level.

***(3) Bar plot visualization:***

***Superstore sales analysis, 2014 - 2017.***

* ***Link to Git Repository:*** [Vasthav\_Dandumenu\_22032815](https://github.com/vasthavd/Vasthav_Dandumenu_22032815)
* ***Link to code file:*** [Question\_3\_Superstore\_Sales\_Analysis\_Bar\_Plot.py](https://github.com/vasthavd/Vasthav_Dandumenu_22032815/blob/ae5b6c6db50e68b5c3d4849ea2fb24a92ecb4c20/Question_3_Superstore_Sales_Analysis_Bar_Plot.py)
* ***Link to dataset:*** [Super\_store\_2014-2017.xlsx](https://public.tableau.com/app/sample-data/sample_-_superstore.xls)
* ***Choice of plot:*** I have chosen the bar plot to visualize the superstore data. Because it is a large set of sales, returns, and location-wise time sales data, I had to figure out what to plot in the first place. It has a lot of columns, containing sales in different categories for example region, state, product categories, and sub-categories over several years of time. For representing sales values for a particular category above I need to plot two variables on the plot showing a relationship which could have been done by a scatter plot as well. But the problem would be with huge data doesn’t give any readable insights in scatter and in the same place bins in the bar plot would give a lively comparison of several values on one axis to each other based upon the other axis sales value.
* ***Code text file and plots:*** Attaching code text file.



Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

* ***Description of plots:*** The above bar plots have been plotted for the sample superstore data of 2014 – 2017, utilising the sales values of several categories of data such as product main category, product sub-category, region of sales, and sales by state. In the four bar plots above. X-axis has been plotted with different type of categories of sales as stated above, Y-axis has been plotted with sales data in dollars as this data is for a fictitious company, the sales values are either too high or low for some categories.
* ***Insights:*** From the above plots we can derive the following insights:
  1. In plot by category technology has the highest sales.
  2. In plot by sub-category chairs and phones have similar sales throughout.
  3. West is the best performing region in sales.
  4. California recorded the highest sales throughout 2014 - 2017.
  5. South region is comparatively the least performing, but the sales are just not very significantly small.
  6. Furniture and office-supplies sales are similar, and the difference is very insignificant.
  7. Similarly east and west regions are very similar in sales throughout.
  8. States of Wyoming, South Dakota, and West Virginia sales very the very least of all barely insignificant when compared to others or can be equated to zero sales.

***(4) Reference Lists:***

* ***Question (1) - Kaggle:*** A Small COVID-19 Dataset <<https://www.kaggle.com/datasets/aditeloo/a-small-covid19-dataset/download?datasetVersionNumber=4>>
* ***Question (2) - Tableau Public:*** Superstore Sales <<https://public.tableau.com/app/sample-data/sample_-_superstore.xls>>
* ***Question (3) - Tableau Public:*** Summer Olympics Medallist Dataset <<https://public.tableau.com/app/sample-data/Summer_Olympic_medallists_1896-2008.xlsx>>