

*Hertfordshire Physics, Mathematics and Computer School*

7PAM2000 Applied Data Science

Assignment 2: Statistics and trends

SEMESTER B, 2023

Module Leader

**Ralf Napiwotzki**

Student ID: 22023513

**Muhammad Faisal Javed**

Dataset link : “ <https://www.kaggle.com/datasets/ashishraut64/internet-users> ”

Github: “”

# Bibliography

Amanda, 2020. *Stack Over Flow.* [Online]   
Available at: https://stackoverflow.com/questions/6077675/why-am-i-seeing-typeerror-string-indices-must-be-integers  
[Accessed 05 03 2023].

deepanshu\_rustagi, 2020. *Geeks for Geeks.* [Online]   
Available at: https://www.geeksforgeeks.org/plot-multiple-lines-in-matplotlib/  
[Accessed 05 03 2023].

government, U., 2018. *UK Gov Data Set.* [Online]   
Available at: https://www.data.gov.uk/dataset/20685d28-c11c-4212-b271-4ded5c8c9486/raw-cows-drinking-milk-quarterly-sampling  
[Accessed 05 March 2023].

jeeteshgavande30, 2021. *Geeks for Geeks.* [Online]   
Available at: https://www.geeksforgeeks.org/plot-a-pie-chart-in-python-using-matplotlib/  
[Accessed 05 03 2023].

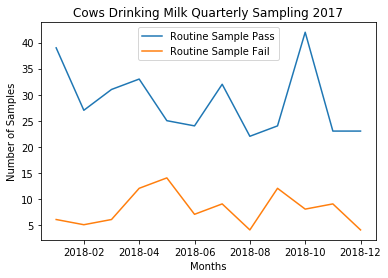
Pandas, 2023. *Pandas PyData.* [Online]   
Available at: https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.plot.bar.html  
[Accessed 06 03 2023].

## Solution:

In this data set I have used data from Kaggle.com and [link](https://www.kaggle.com/datasets/ashishraut64/internet-users) in provided on the end of this document.

Study in *Raw Cows' Drinking Milk Quarterly Sampling 2018* shows the number of samples taken by the government authority and put the samples in testing, then taken out the results of the samples and divided into two major parts; passed ones and failed ones. In this document we are going to visualize the samples using three different types of techniques.

**Line Plot**

 (Amanda, 2020) (jeeteshgavande30, 2021) (Pandas, 2023) (deepanshu\_rustagi, 2020)

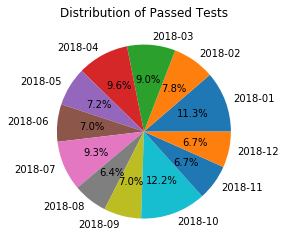
**Reason for Choosing Line Plot:**

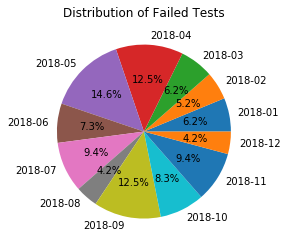
Reason for choosing line plot is because we can easily pointout the difference between given data just by looking at the line movements

**Explanation:**

In the figure shown above, we can see that the number of tests taken from the government authority in the whole year of 2018 on monthly basis. And there is big gap between passed and failed test, which is good thing in it. Moreover, we can see that in the months where passed tests are increasing, failed tests are decreasing. And where passed tests are decreasing, failed tests are increasing. (government, 2018)

**Pie Chart**





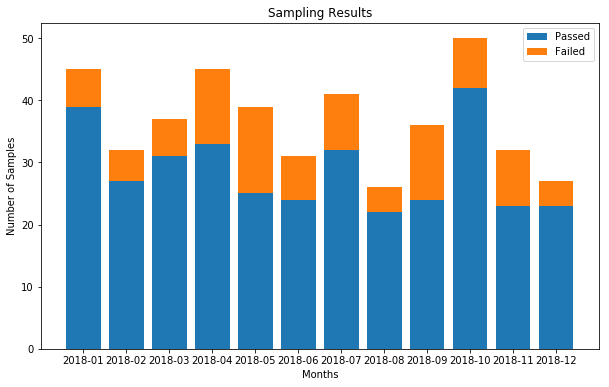
**Reason for Choosing Plot Chart:**

Pie chart is being used for the visualization of data because in pie cart we can see the difference of the outcomes by giving then different colors.

**Explanations:**

In pie chart diagram, there is number of distribution of failed and passed test taken and out of which how many are failed and passed per month in whole year.

**Bar Chart**



**Reason for Choosing Bar Chart:**

Bar chart is the technique I chose because bar chart is good for the visualizing the difference of data at the same time and comparing at the same time.

**Explanations:**

In the diagram we can clearly see the total number of tests taken and from which how many have passed and how many have failed the test. As shown in the diagram there is huge difference between passed and failed tests. As blue color represents the passed tests or orange for the failed ones, most of the tests are passed and only few of them failed each month of the year. Numbers are varying each year, but most of them are passed tests.