DATA VISUALIZATION

SOUMONOS MUKHERJEE | *VATSALYA NAYAK* | *SAHIL SHIRODKAR*

Case Study : A Bank Marketing campaign in Portugal and Economic indicator of the Nation

We chose the country Portugal since it has a large population but is less talked about in recent times as compared to other European nations. To understand the economic standing of Portugal we decided to analyse the data set of a Portuguese banking institution.

Also the reason behind choosing the World bank Data was to understand the country's economic movement over a timeframe. The reason behind choosing a specific indicator is that, By researching through all, we have found this indicator transition to be highly correlated to the GDP and Growth coefficient evaluation of Portugal.

The StrongHold and Edges of Our project (Standing out from the crowd):

1. We have left no stones unturned. Analyzed and plotted every single feature available to us.

2. We plotted every possible interaction between and amongst the variables.

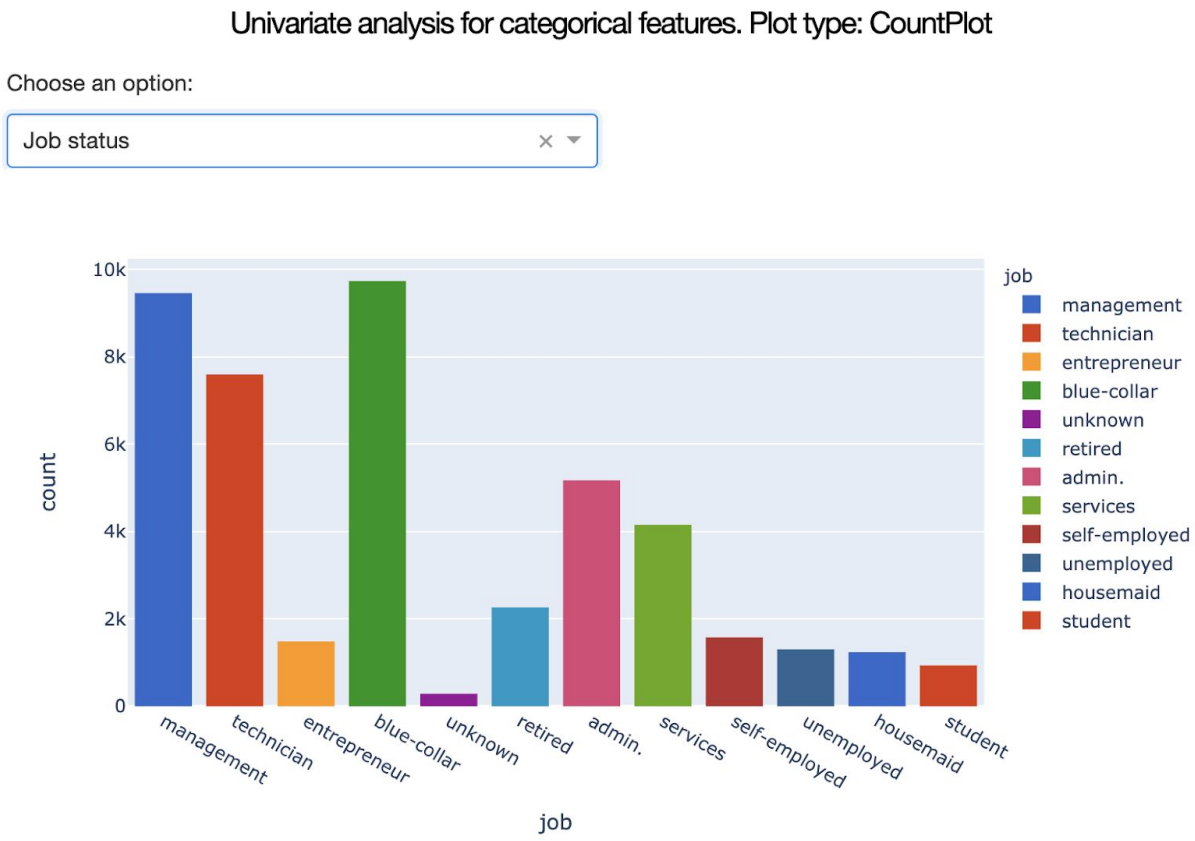
3. We have implemented the Dashboard with Interactive Plotly charts and we have used multiple dropdowns. This has enabled us to precisely describe every kind of distribution and interaction without populating the dashboard much.

4. The Dashboard is user friendly and One can group by plots with any possible combination to dive deep inside the analytics.

5. 3D visualization is novel and we have tried to implement it conveniently .

6. The codes are clean, neat and comprehensive. Executing the code is hasslefree. One just needs to get into the app directory and can implement the .py file into a console. You will get the URL. Just paste the same in your browser et Voila!

The First Graph is a Count plot for the categorical univariate distribution



Reasons behind choosing Count plot:

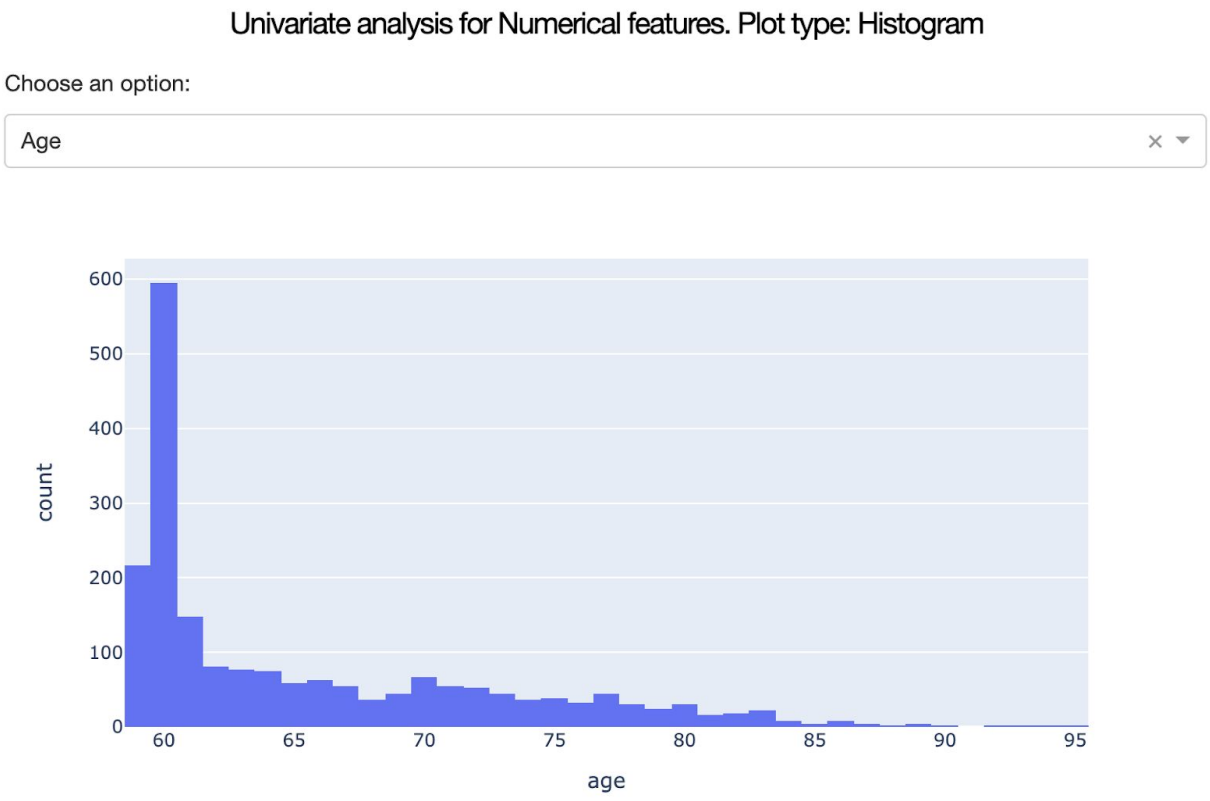
1. The variables to be analyzed are categorical, We can not have a very comprehensive density distribution.
2. Count plots are easy to comprehend. Upon moving the cursor on specific class bars, they show us the exact number of samples belonging to that specific class.

Attributes in the count plot

Job Status	Type of Job
Marital status	Marital Status
Education level	Level of Education of the

Defaulter's- Status	Defaulter's
Loan History	Loan History of the client
Contact-method	How the client was contacted
Contact-month	Month of contact
Out-come in the previous campaign	Success or failure
Subscription status	Subscription Status of the clients

The second Graph is a Histogram which will plot the numerical attributes with its Distribution



Reasons behind choosing the Histogram plot type:

1. The features to be analyzed are essentially continuously distributed. They can not be shown in individual categorical counts.

2. Histograms are efficient in showing the overall distribution the feature is following. It is also useful when we want to see how many samples lie in the extreme ends and where is the possible highest density observable of the distribution.

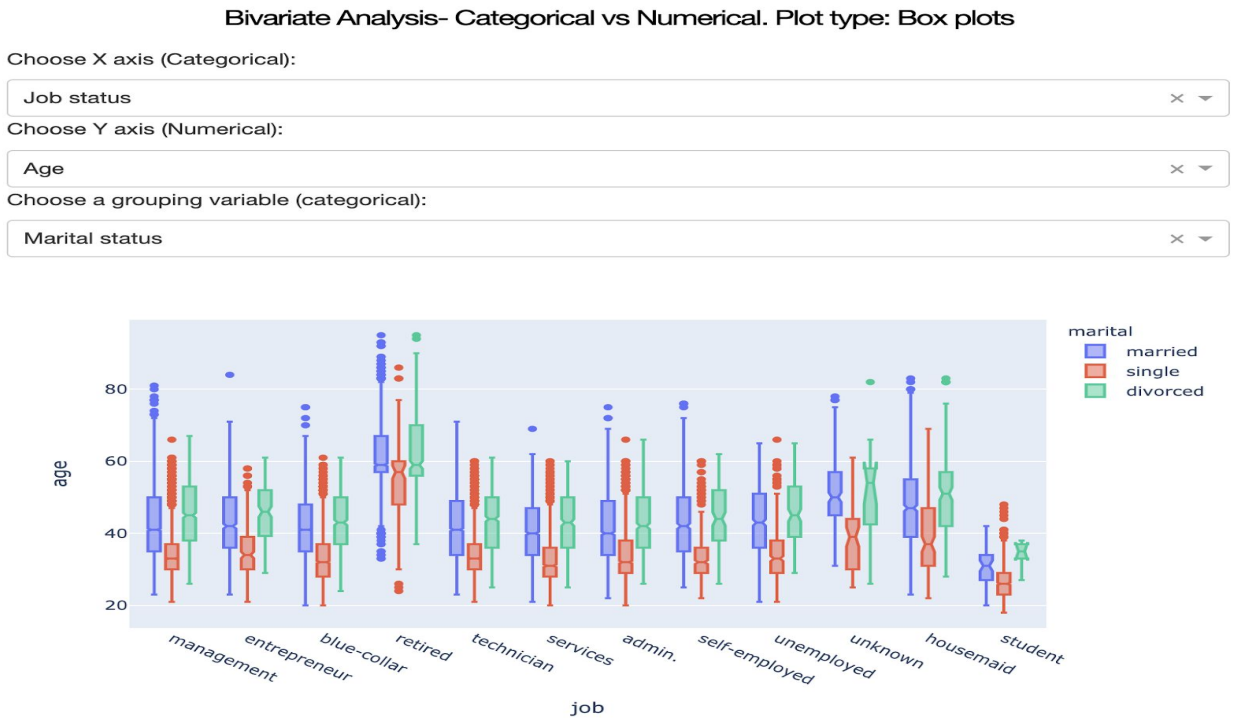
Attributes in the Histogram

Age	Age of the client
Bank Balance	Bank Balance of the client
Total number of days contacted	Numbers of days the client was contacted for the campaign
Last contact Duration(Secs)	The duration in seconds the client was contacted for the campaign
Number of times clients contacted during the campaign	number of contacts performed during this campaign and for the client
Number of times client contacted previously	Number of times the client was contacted before the campaign

Quick Analytics

From the histogram, we can see that the Bank balance of most of the people lies between 6k to 7k. Mostly the age of the clients is above 55 years of age.

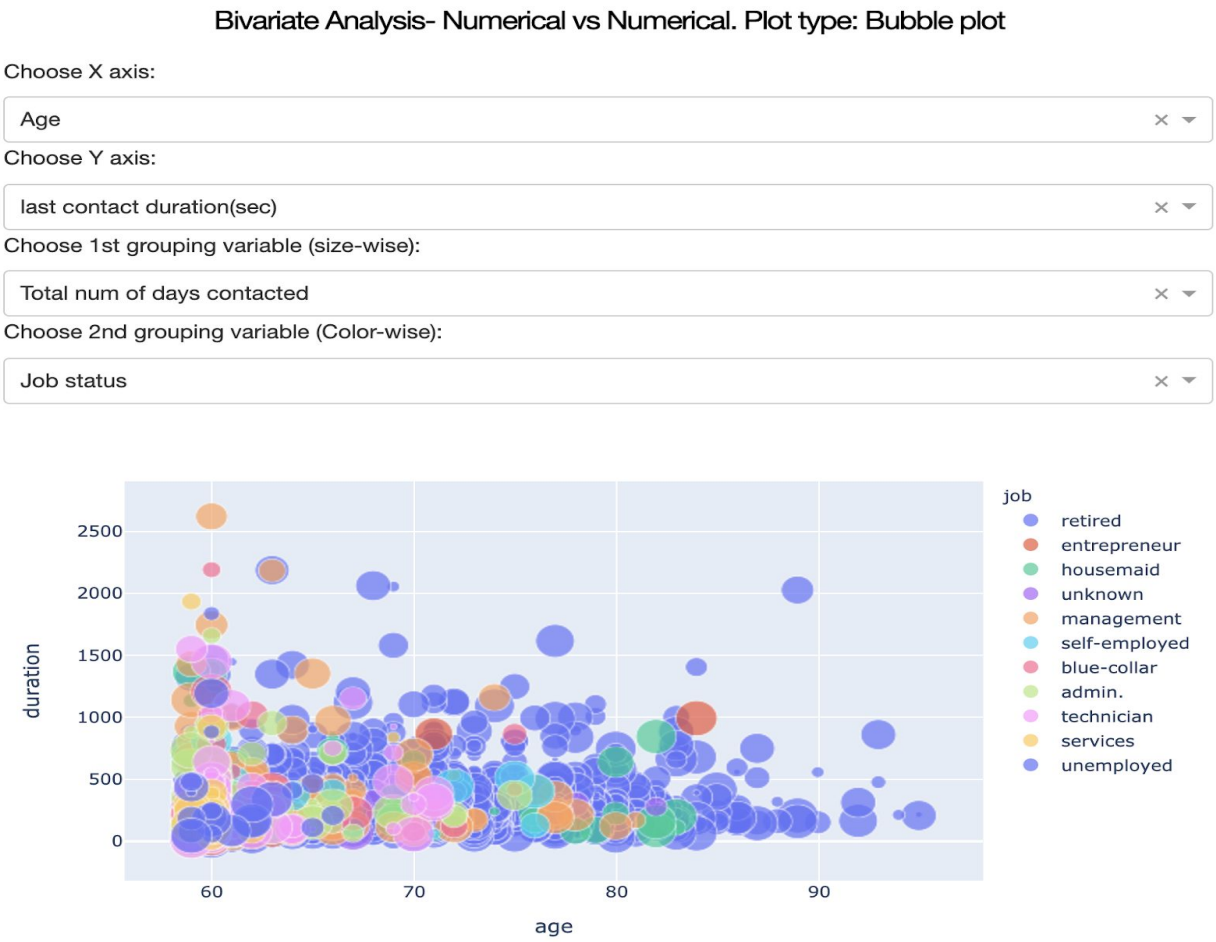
The third graph is Box Plot Representation:



Reasons behind choosing the plot type:

1. Boxplots are extremely efficient when it comes to comparing a numerical feature with a categorical counterpart as we do not get the distribution type of the numerical feature over the specific classes in Barcharts.
2. The interactive chart, if one moves the cursor to a particular box, they can know the exact mean, median and quartiles of the numerical feature on the specific categorical class and their numbers grouped by respective classes.
3. Our integration of the grouping variables and the dropdown enables one to exactly analyze all the variables with each other and group them accordingly onto specific classes.

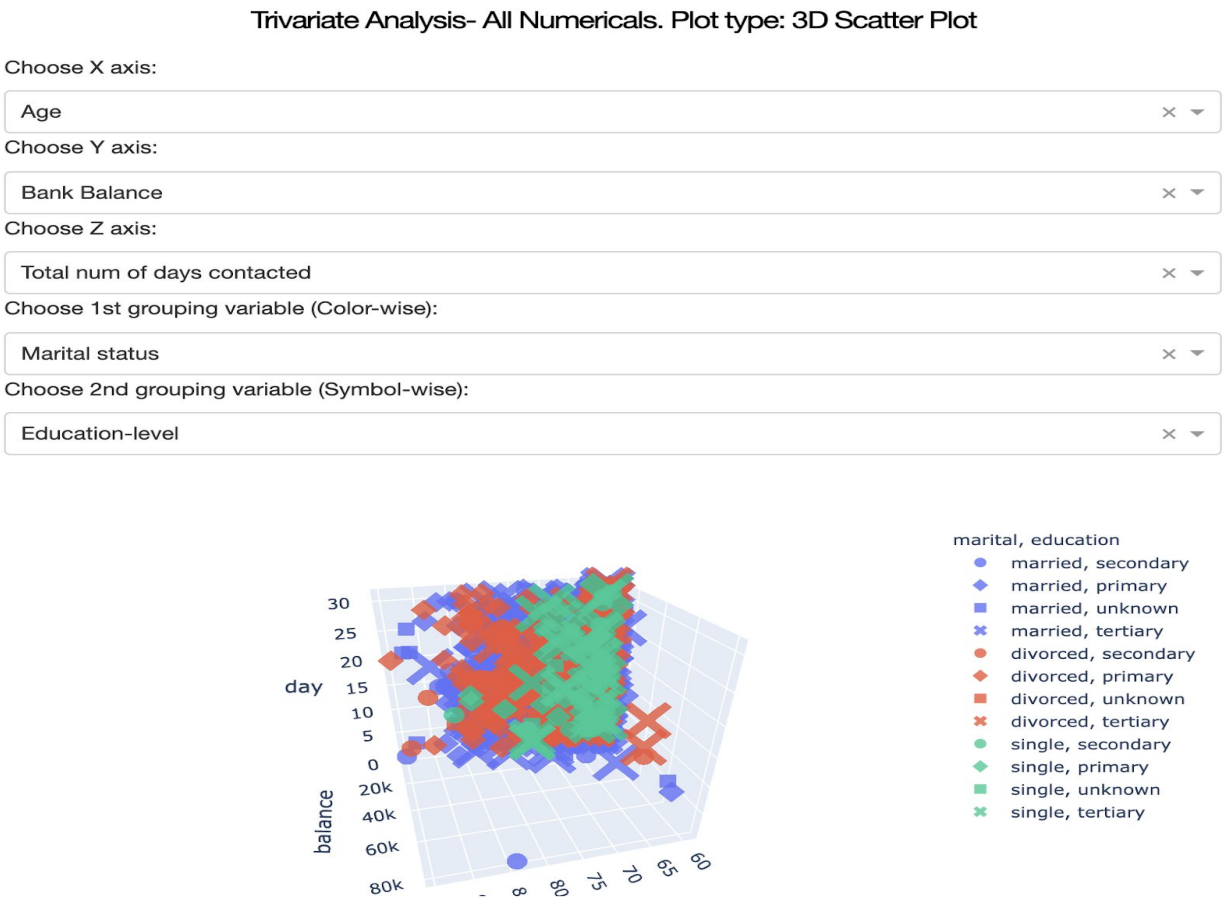
The Fourth Graph is a Bubble plot



Reasons behind choosing the Chart type: Bubble plot

1. Bubble plots outperform the traditional scatterplots in numerical bivariate analysis as we get another option to group the scattered data by another categorical variable to go deep dived into analyzing the distribution.
2. Specific color groups and Size groups enable us to analyze 2D data with discrimination over classes.

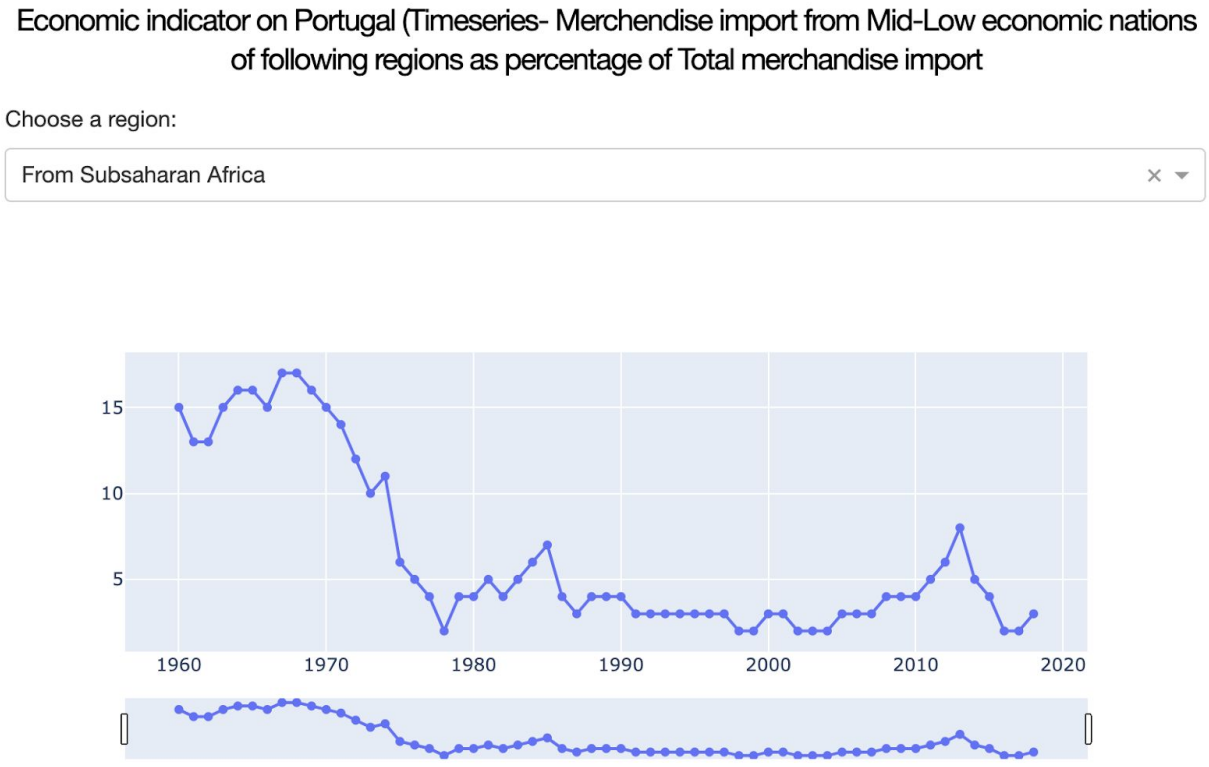
The fifth Graph is a 3D scatter plot



Reason behind choosing the Graph type: 3D Scatter plot

1. This is a novel way to visualize data in 3D . It is unprecedented in PyData Stack.
2. Primarily it looks messy but we can zoom in inside and see the distribution in 3D.
3. The grouping features let us discriminate the classes over the distribution.
4. Moving axes gives us the real figure with specific component analysis.

The sixth graph is an interactive Economic indicator on Portugal (Timeseries- Merchandise import from Mid-Low economic nations of the following regions as a percentage of Total merchandise import



Quick analytics:
From this chart, we can conclude that in recent years the merchandise import has been quite less overall but it has increased from the European and Asian countries as compared to African countries.

Overall Analytical Conclusions:

- Portugal, in general, comes out to be quite educated and overall a low saving populated nation.
- The usual pattern is seen when it comes to the Job types and labour distribution of the country.
- Management people are more prone to subscribe to the plan.

- The bank tried to target most of the people in the age range of 50-60. Their job status is mostly retired and that can affect their affinity to subscribe to the plans. It would be great if they would focus more on the younger working population.
- The 3D- Scatter plot gives us the flexibility to analyze a total of 3 numerical variables at once, zoom in inside to see how the data points are distributed in 3D space. And the categorical grouping feature dropdowns enable us to group the distribution with categorical hues.
- The time series Data gives us the impression that Portugal is not doing very well when it comes to the trade scenarios with African Countries (Specially the Sub Saharan areas). But its trade and import has substantially increased with Asia and Europe.