Data Visualization CA1

Group Report

Himansh   Shruthi   Nikhil

Dataset

Source

The dataset is taken from the UCI Irvine machine learning repository ,dataset is about the phone campaigns with the individual customers of Portuguese banking services, Data is Multivariate, Business related and classification related problem. Data cleaning is done in python pandas library checked for missing values and Info of the dataset variables everything looks fine no missing values, identified unique categorical values in each variable this helps how we are going to deal with this data.

The dataset can be accessed at: https://archive.ics.uci.edu/dataset/222/bank+marketing

Variables

1. **Age:** Indicates the age of the individual contacted during the campaign. (numeric)

2. **Job:** Indicates the type of job that the individual holds. (Categorical: Retired, Student, Management, Housemaid, Self-employed, admin, technician, unemployed, unknown, services, entrepreneur, blue-collar)

3. **Marital:** Indicates the marital status of the individual (Categorical: Single, Divorced, Married, Unknown)

4. **Education:** Indicates the level of education achieved by the individual. (Categorical: Primary, Secondary, Tertiary, Unknown)

5. **Default:** States whether the person has a default loan. (Binary: yes/no)

6. **Housing:** Indicates whether the individual has housing loan. (Binary: yes/no)

7. **Loan:** Indicates whether the individual has taken a personal loan. (Binary: yes/no)

8. **Balance:** Indicates the bank balance of the individual. (numeric)

9. **Contact:** Indicates the mode of communication used of contacting the individual during the campaign. (Categorical: Cellular, Telephone, Unknown)

10. **Month:** Indicates the month when the individual was last contacted. (Categorical: Jan, Feb, Mar, Apr, May, June, July, Aug, Sep, Oct, Nov, Dec)

11. **Duration:** Indicates the duration of the last call made with an individual. (numeric in seconds)

12. **Campaign:** Indicates the number of times an individual was contacted during the current campaign. (numeric)

13. **Pdays:** Indicates the number of days it has been since the person was last contacted. (numeric)

14. **Previous:** Indicates the number of times the individual was contacted during the last campaign. (numeric)

15. **Poutcome:** What was the outcome of the previous campaign for the given individual. (Categorical: Success, Failure, Other, Unknown)

16. **Y:** Indicates the outcome of the current campaign. (Binary: yes/no)

Vision

The goal of the project is to analyze whether customers will subscribe to a term deposit with the bank or not. It is based on 3 main factors throughout the report:

1. Demographic Characteristics: These characteristics consist of the personal details of the client including age, education, marital status, and job type.

2. Financial Characteristics: These attributes mainly encompass the financial conditions of the individual including their loan status (personal, default, or housing), and their bank balance.

3. Campaign Characteristics: These terms mainly include details from the current and previous campaigns conducted to market the term deposit idea to different individuals.

Our goal is to understand the changes in purchase rate, i.e., the rate at which customers invest in the term deposit for the bank, based on these different characteristics, and find the key drivers that affect purchase rate, while also identifying high-value customers and how more customers can be convinced to invest with the bank.

Power BI

Key Insights

- Most number of purchasers of the term deposits are either retired or students, could be categorized as a naive group.
- Majority (64%) of the customers who purchased the term deposit in the last campaign have purchased it in this campaign too.

- Majority of the customers contacted for this campaign were not contacted in the last campaign - meaning the bank has gained a lot more customers since then.

At-Risk Groups

- The report shows that for both "failure", and "success" as the outcome of the previous campaign, the lowest purchase rate groups are the ones with default loans.
- Another group at-risk of not investing with the bank are the ones that have loans, or housing loans, or have lower than the average balance of all customers, for most of the previous outcome cases.
- If possible, the bank should look into the loan status of individuals as a key factor in marketing their proposals.

High-value customers

- The most active investors are the ones who were contacted the most number of times, especially via the telephone communications mode. Higher communication frequency with customers could be a key identifier in increasing investments with the bank.
- Those who invested in the previous campaign seem to have a higher average purchase rate for the current campaign where higher number of contacts made, steeply increases purchase rate. These are high-value customers for the bank.
- The people last contacted in March, September, October, and December via cellular and telephone communication have a higher purchase rate.
- Cellular communications mostly had short durations for this campaign and had a lower purchase rate, whereas telephone and unknown contacts showed higher purchase rate and slightly longer durations.

Description of Excel

The data we have and going to lead with more categorical data end objective will a individual take term deposit Yes/No in the bank, many factors contribute to target variable like campaigns and job categories, marital status , average call duration with the each age categories using this features we are going to use the Excel based technology.

Pivot tables & Charts brief understanding

- Stacked bar chart(Job vs Deposit subscription) is taken to explain the count of the subscription based on the job categories This information is helpful to the banking organisation to make campaign decisions based on the job categories features taken inn the pivot table are the job variable to the row labels and the y output variable to

the values label(count) column label is y variable itself. By this visualisation highest subscriptions are from the management job category but with the highest count of rejections.

- Clustered column chart(Age categories vs Deposit Subscription) This chart explains which age categories perform well to take the term deposit and the ratio of subscribed and not subscribed clearly here middle age categories not subscribed ratio is high it make slight unbalance, balanced group to target is the old age category it as the balanced no of no's and yes for subscription.
- Line chart(Month vs campaign success) with this chart we came to know that campaign success is with respect to month, in the month of may there are highest rejections to the subscriptions due to various reasons, in the month of Dec it has the highest no of subscriptions.
- Donut chart(Total call duration by age categories) gives us information regarding average percent of duration of time sent with the different age categories Adults, and old people spent max time, one more Donut tells us total percent of the unsubscribed and subscribed.
- Bar chart(Contact vs subscription) by visualization we can conclude that communication by cellular, telephone don't make much difference. They are exactly the same conclusion communication medium don't make impact on the subscription.

Features used to make viz interactive

Excel slicer used to make the dashboard interactive balanced, connecting all the slicers with option Report connections with all the pivot charts selecting the slicer option this changes all the data in the entire dashboard pivot charts with respect to selected specific option.

Group Appraisal

Power BI (Himansh)

Strengths:

1. Good analysis. Provides good insights about all aspects.
2. Good use of the number cards which are interactive with all the plots.
3. The text provides more insights which is a positive point.

Weakness:

1. The scatter plot on the last page is too clustered.

Tableau (Shruthi)

Strengths:

1. Good plot overlapping the jobs with education.

2. Good line chart clearly showing married people have the most balance.
3. The plot between balance and job shows which jobs have the highest pay.

Weakness:

1. The plots do not involve the target variable.
2. The pages were not included. Not all the plots are interactive. Only the education ones are, with the slicer.

Excel (Nikhil)

Strengths:

1. Color declared for the y variable, which makes it easier to analyse.
2. Good use of slicers
3. The bar charts and line chart give good insight.

Weakness:

1. Should have used different colors for marital status because they overlap with the colors of the y variable.
2. The pie chart of Average duration of calls does not say what it means.