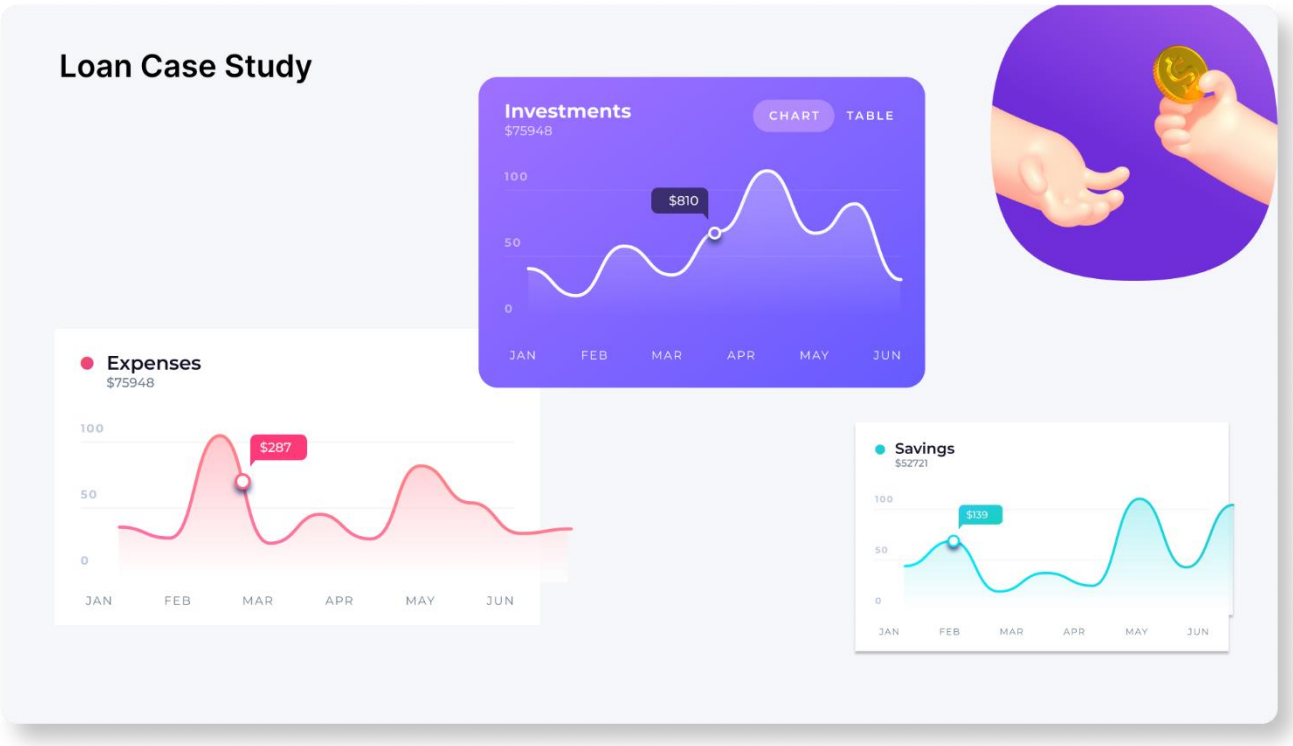
Bank Loan Case Study

AUGUST 11

By Aastha Kumar

Description

This case study aims to give you an idea of applying EDA in a real business scenario. In this case study, apart from applying the techniques that you have learnt in the EDA module, you will also develop a basic understanding of risk analytics in banking and financial services and understand how data is used to minimize the risk of losing money while lending to customers.



Approach

First, I downloaded the dataset from Google Sheets onto my personal device for making modifications. I then used my knowledge in statistics and used different formulas in excel to draw necessary conclusions about the company.

Tech-Stack Used

- Google Sheet

Insights

A. Identify the missing data and use appropriate method to deal with it. (Remove columns/or replace it with an appropriate value)

B. Identify if there are outliers in the dataset. Also, mention why do you think it is an outlier. Again, remember that for this exercise, it is not necessary to remove any data points.

C. Identify if there is data imbalance in the data. Find the ratio of data imbalance.

   Hint: Since there are a lot of columns, you can run your analysis in loops for the appropriate columns and find the insights.

D. Explain the results of univariate, segmented univariate, bivariate analysis, etc. in business terms.

E. Find the top 10 correlation for the Client with payment difficulties and all other cases (Target variable). Note that you have to find the top correlation by segmenting the data frame w.r.t to the target variable and then find the top correlation for each of the segmented data and find if any insight is there. Say, there are 5+1(target) variables in a dataset: Var1, Var2, Var3, Var4, Var5, Target. And if you have to find top 3 correlation, it can be: Var1 & Var2, Var2 & Var3, Var1 & Var3. Target variable will not feature in this correlation as it is a categorical variable and not a continuous variable which is increasing or decreasing.

F. Include visualizations and summarize the most important results in the presentation. You are free to choose the graphs which explain the numerical/categorical variables. Insights should explain why the variable is

important for differentiating the clients with payment difficulties with all other cases.

Resources

Loan Case Study

Result

It would have been impossible or very time consuming for an ordinary human to be able to process such large bits of information. A computer can do it within seconds with the right commands. That's why companies like Instagram hire data analysts to control the waves of data they collect every day, makes sense of it, and then draw conclusions or make predictions. This is the process of turning data into insights, and it's how analysts help businesses put all their data to good use.

The more detailed definition you learned earlier is that data analysis is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

Data analytics can help organizations completely rethink something they do or point them in a totally new direction. For example, maybe data leads them to a new product or unique service, or maybe it helps them find a new way to deliver an incredible customer experience.