Credit Car Prediction Dataset

# Final Report by Ahmet TURK Introduction

# This report documents the execution and analysis of the project using the dataset provided in the “test\_combined” worksheet from the “test\_data.xlsx” file. The primary objective of this project is to analyze the given data to identify patterns and insights related to financial status, employment, and other demographic factors of individuals. This analysis aims to uncover any underlying issues and propose actionable solutions based on the data.

# Business Impact

By analyzing this dataset, we can obtain valuable insights into the specific channels credit cardholders use for their purchases, the types of products they buy with their credit cards, and the product categories they prefer. This detailed analysis enables us to customize privileges such as coupons, discounts, and other perks to match individual cardholders' purchasing behaviors. By doing so, we can more effectively cater to their preferences, thereby enhancing their overall satisfaction and loyalty.

This information will serve as a crucial guide for the company in determining the types of benefits and rewards that should be offered to cardholders. Tailoring these advantages to align with cardholders' purchasing habits ensures that they feel valued and understood, which can significantly boost their engagement and loyalty to the company. Ultimately, this approach not only improves customer satisfaction but also strengthens the relationship between cardholders and the company, fostering long-term loyalty and sustained growth.

# Data

File Name: test\_data.csv

Description: Credit Car Prediction Dataset Unlocking Financial Access: Predicting Credit Card Eligibility

Dataset Details: 7293 Rows & 16 Columns

Size: 960,14KB (0,9MB)

Source: [Kaggle](https://www.kaggle.com/datasets/tanayatipre/car-price-prediction-dataset/data)

# Data Analysis & Computation

## Data Profiling:

* Apply Excel filter to the dataset.
* Examine dataset through Excel.
* Apply Excel IF function to the dataset.
* Apply Min(), Max(), Median(), and Avg() functions to the dataset
* Apply Excel filter to the dataset.
* Notable Features (Prior to Export )
  + 18 Unique Job Title Category
  + Revenue – Examine the revenue of the each customer.
  + Account Age – Using IF function to convert from minus values to accurate values of the account age.

### Observations:

* Some records consist of minus values instead of string fields such as age, employment length, account age etc.
* Job title holding blanks values.

# Data Set & Wrangling:

The dataset contains various attributes of individuals, including their gender, car ownership status, property ownership status, number of children, income, employment status, education level, marital status, and dwelling type. The initial inspection revealed several columns with missing or irrelevant data, such as unnamed columns and those containing summary information. The following steps were undertaken to clean the data:

* Removed unnamed columns that do not contribute to individual records.
* Handled missing values by either filling them with appropriate values or removing the rows if the missing data was substantial.
* Ensured consistency in categorical data by standardizing the entries.
* Using Excel, sorted attributes by field types. Some attributes contain minus value fields when they should be deleted. 2 columns were deleted.
* Filtered “Job Title” for any records that belong to all job title.
* Filtered “Marital Status” for any records that belong to all status of marital.
* Filtered “Gender” for any records that pertain to male and female.
* Using aggregate functions in Excel, separated the income of the customer revenue with gender.
* Using filter “family member count” for any records that the number of the family member.

# Data Analysis

## A pie chart of gender Description automatically generatedGender Distribution

The provided chart illustrates the distribution of credit card usage among different genders. It specifically shows the percentage of males and females who use credit cards. This data indicates that a significantly higher percentage of females (62%) use credit cards than males (38%).

## Total Records & Blank (Null) Data

The dataset includes 7293 rows and 18 columns, 9 of which contain numerical data and 9 of which contain textual data. While checking the blank (null) data, the percentage of the blank number is the highest, at 31%, in the job title.

## Credit Card Holders by Gender and Income

## A graph with numbers and a bar Description automatically generatedThe bar chart illustrates credit card holders' total income, segmented by gender. The comparison includes male and female categories, showcasing the total income generated from each group.

Male Credit Card Holders:

* Total Income: 5,200,298,775

Female Credit Card Holders:

* Total Income: 8,353,139,355

## Income and Employment Status

When looking at the income and employment status of credit card holders, exactly 50% are classified as “working”, following by 27% as “commercial associate”, and 14% as “pensioner”, respectively.

Considering income, certain values are as follows:

Media 157500

Mean 185867,226

Minimum 27000

Maximum 1575000

Sum 1355343813

# Conclusion & Future Work

The analysis provided several actionable insights, such as the impact of employment status on income and the characteristics of high-risk individuals. Future work could expand the model to include more advanced algorithms like random forests or gradient boosting machines for potentially better performance. Additionally, incorporating more granular data (e.g., transaction history) could enhance the predictive power of the models.