

Supervisor

Project Coach

Selim Balcısoy

Ahmet Bozkurt

Project Team

Abdulsamet AKTAŞ
Oğuzhan BERBEROĞLU
Oğuzhan TAVŞAN

ABSTRACT



Agency for the insurance needs individuals; advise them, negotiate with insurance companies on behalf of them, act on behalf of the insured in determining the damage and on behalf of the insurance company in the branches it is authorized. Agency earns a commission from the policy premium it directs or deducts from the insurance company. Doğuş Teknoloji works with VDF Sigorta and it works as agency company. The project, goes under industry-oriented projects within this company. The project scope is data analysis on policy agreements for future strategy estimation within this company. Deciding whether the insurance company will be profitable for the next year.

OBJECTIVES

Estimating Future Sales

Classification of Products

Estimating Product Income

- Estimating future monthly and annually expected profits based on the previous months and years.

- Be able to classify the dataset according to profits. Starting with the most profitable products to less in six different categories.

- Given the inputs such as product type, insurance company code, agency code and other features, be able to estimate the income of this product.

PROJECT DETAILS

Size

- Entries : 4.454.227
- Features : 41

Data Types

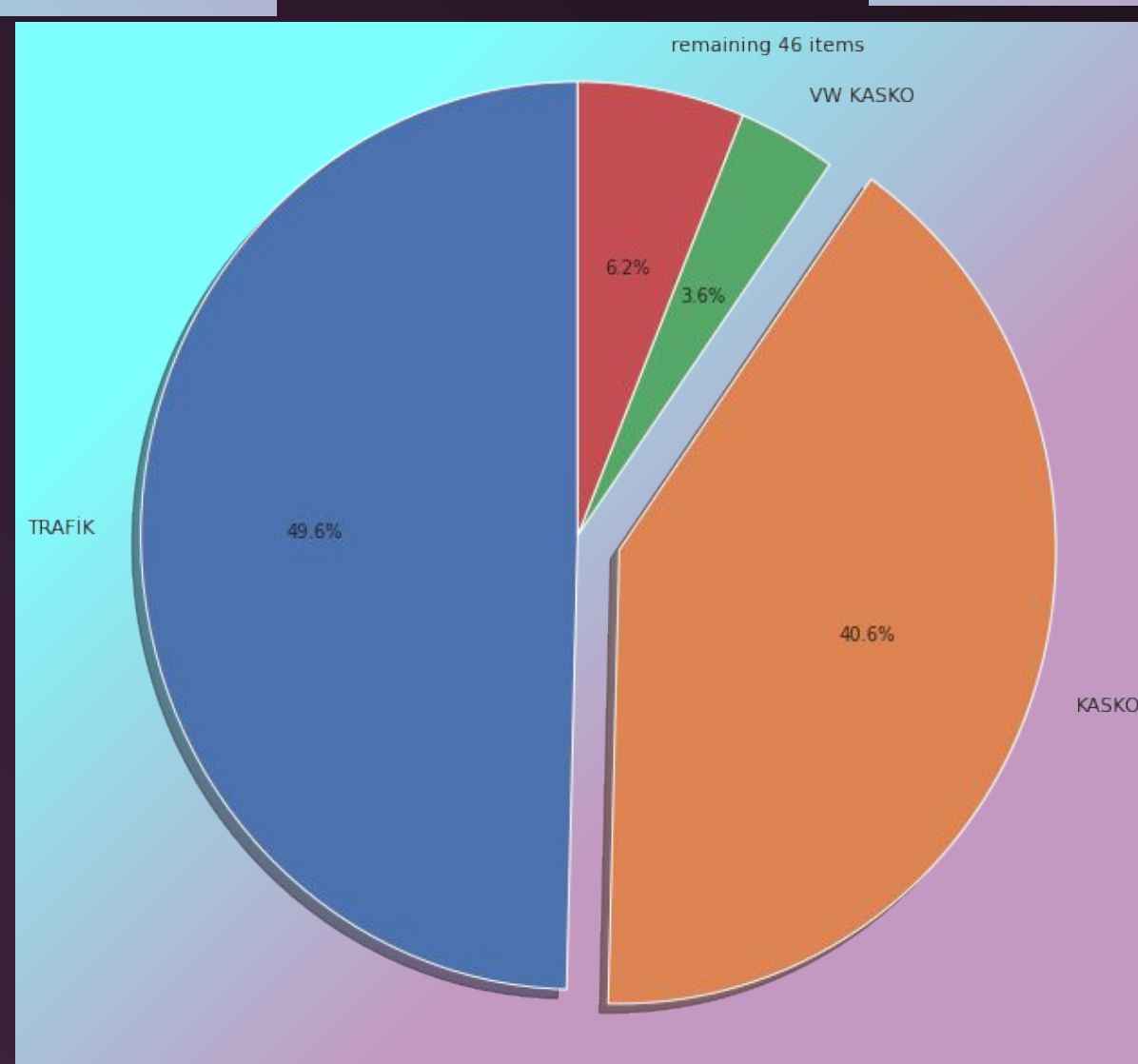
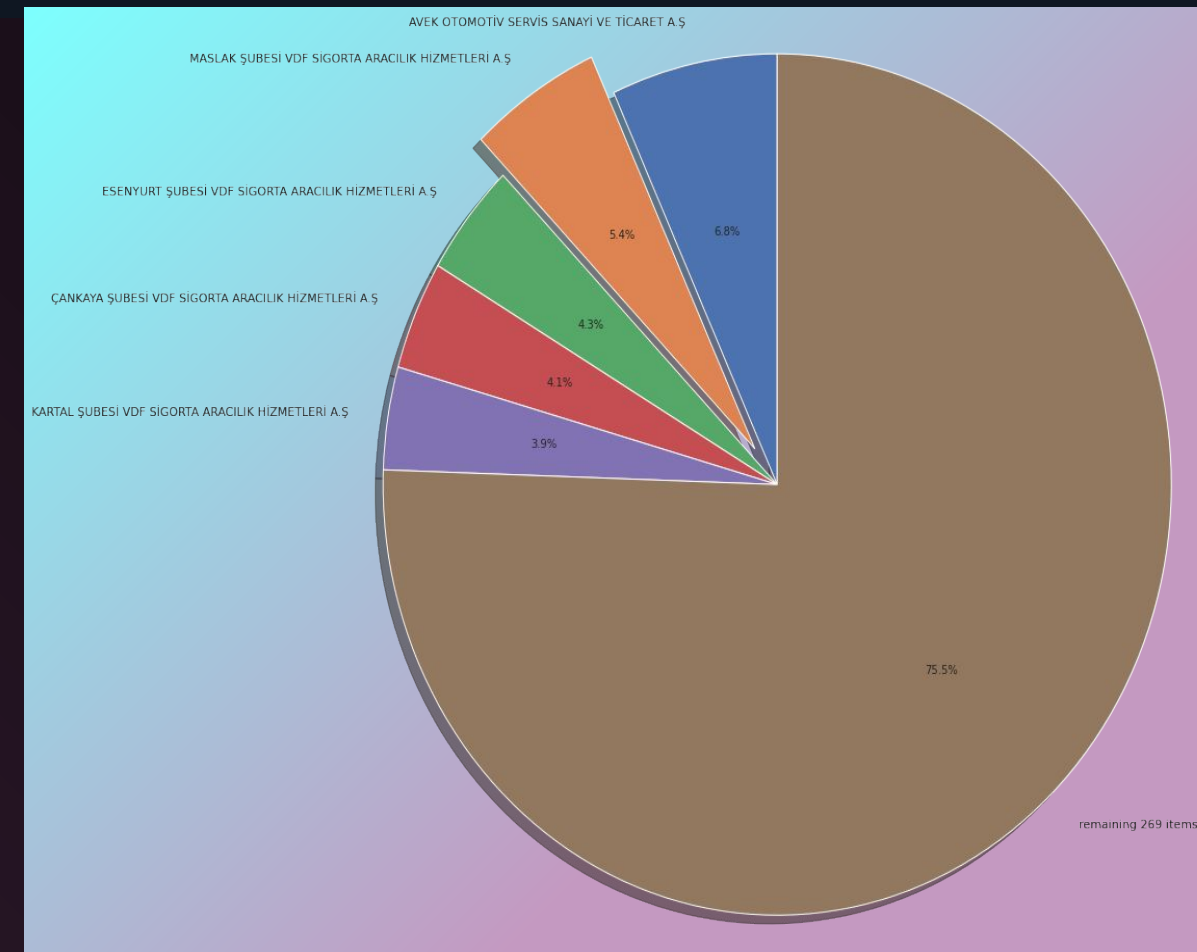
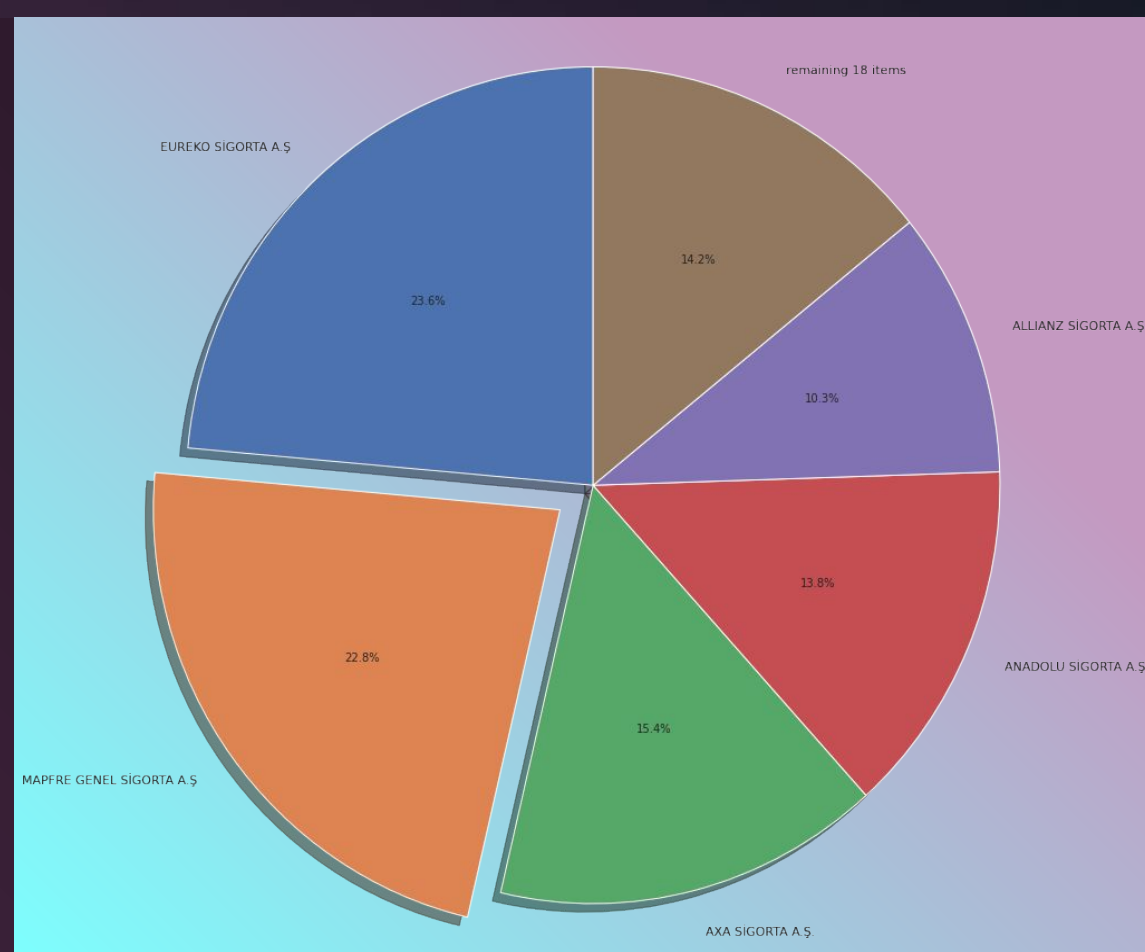
- Bool : 1
- Int : 6
- Float : 3
- Object : 31

Info

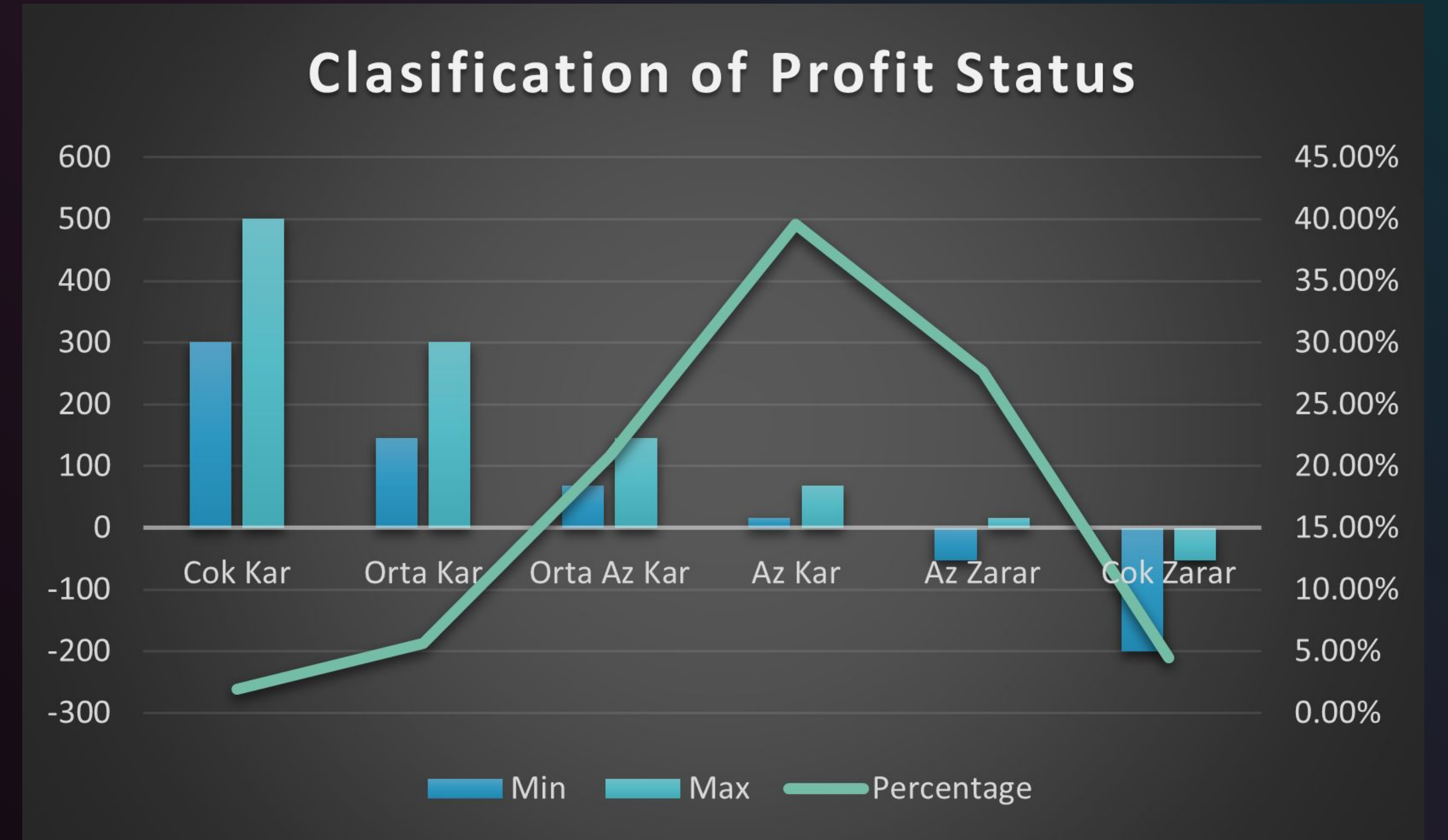
- Data includes entries starting from 2004 to 2021

Details

- 25 Different Insurance Companies
- 293 Different Subordinates
- 52 Different Product Categories



PROJECT DETAILS CONT.



XGBoostRegressor

LGBMRegressor

KMeans Clustering

- XGBoost regressor used during the estimation of the product profit with different hyper parameters. Our aim was to get better results in terms of RMSE.

- LightGBM regressor is another algorithm that estimates the profit of the product.

- KMeans is most common clustering algorithm. We have used this algorithm to determine the classification of the products in terms of profit.

Results:

Mean Absolute Error
MAE_XGB: 4.4746

Mean Squared Error
MSE_XGB: 121.2836

Root Mean Squared Error:
11.0128

Results:

Mean Absolute Error
MAE_XGB: 2.9104

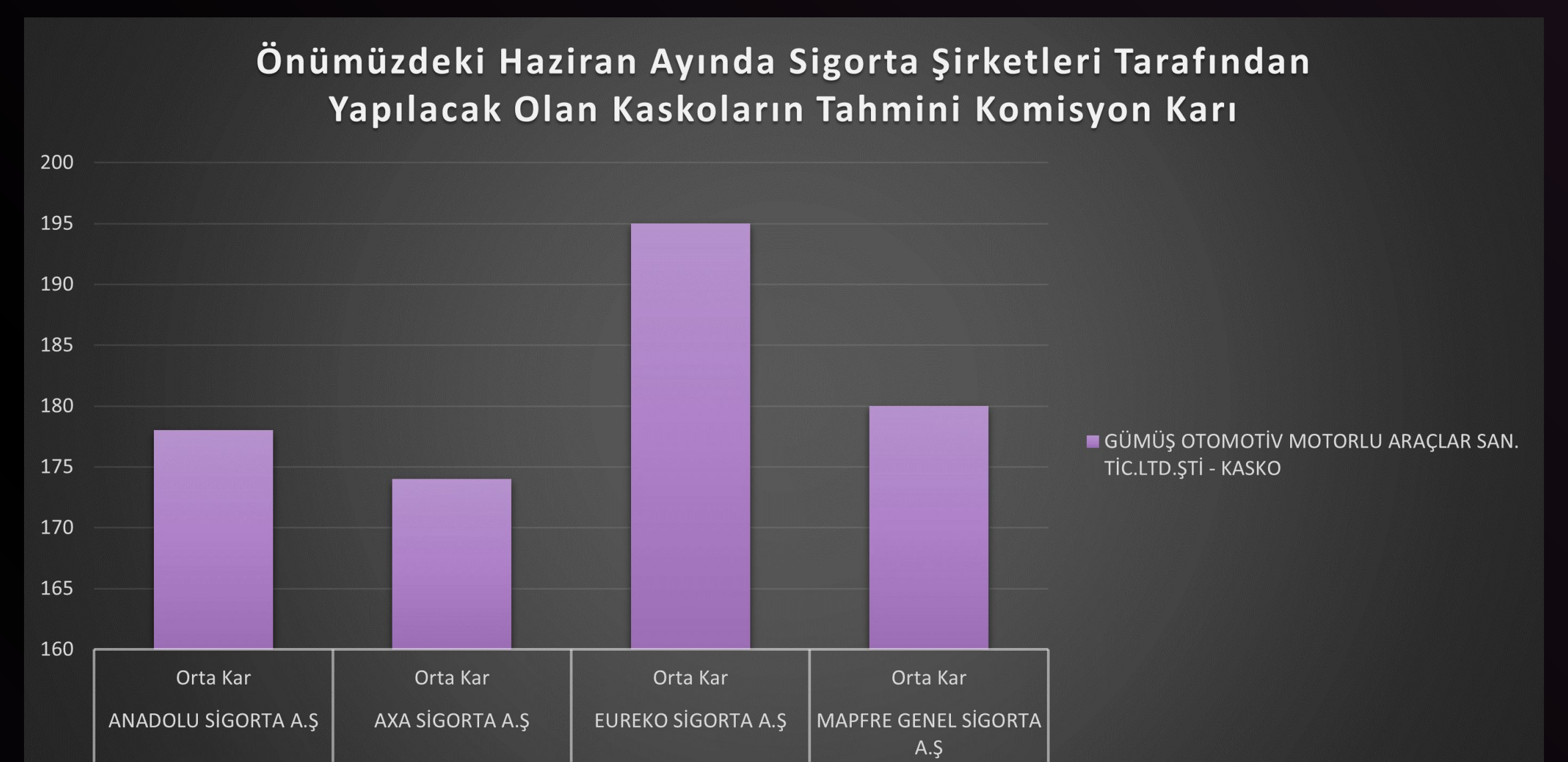
Mean Squared Error
MSE_XGB: 70.4437

Root Mean Squared Error:
8.3930

Results:

We have labelled classes starting from most profitable to less profitable in six categories

CONCLUSION & FUTURE WORK



Categorized The Dataset Clustering

We have labelled the data according to profit of the products

- Divided the dataset into 6 categories

Predicted Product Income Regression Models

Developed two models for estimating the profit.

- Chose best model for digital assistant
- Combined clustering and regression in the digital assistant

Future Work

Model Improvements

- An Application can be built based on the digital assistant for future sales of the company

REFERENCES

Insurance Information Institute. Insurance Handbook A guide to insurance: what it does and how it works. New York. Retrieved from https://www.iii.org/sites/default/files/docs/pdf/Insurance_Handbook_20103.pdf