

# **Product Based Monthly Prediction and Tracking of Plastic Packaging Weight**

## **Public Project Summary**

Ersan Kucukoglu

Central European University

Department of Economics and Business

MSc in Business Analytics

June 2022

## Background

The purpose of this Capstone Project is to assist the client, a company in the retail industry, in creating a data product with a business-driven data analytics approach. The project aims to build a data product, which allows the user to see the sold plastic packaging in tones and compare it with the business goals also by seeing the predicted outcome till the end of the business year. To reduce the negative impact on the environment, the business goal is to reduce the plastic packaging sold with our product by 20% by 2025. By developing a data product that calculates the KPI of plastic packaging sold by suppliers contracted with HU in tons and compares it with targets, immediate response in case of a risk will be possible.

Key problems:

- evaluating the business goals is done quarterly
- there is only a high effort possible to follow the measures within the year, no forecast is available

In order to meet the project's requirements, I had to answer following questions.

What is the sold plastic packaging quantity in tones? What is the estimation of sold plastic packaging till the end of the business year? Which items are significantly high and should be optimized (packaging change, like packaging change) based on the past / predicted values? This process should be automated and implemented with a dashboard for visualization.

## Project Summary

In this project, it was aimed to reduce the plastic packaging sold (in t), and in this context, the client aimed for a data product as project output that allows the user to see the plastic packaging sold in tons and compares it with the business targets by seeing the predicted result until the end of the business year. In order to achieve these goals, Time Series Analysis was performed according to the data provided by the client, using tools such as Databricks, Python, and SQL. After data cleaning, data exploration, and time series analysis, forecasting models were built to predict the plastic packaging weight for each item. With the Prophet model with the best performance, monthly predicted values were created until March 2023, at the end of the business year. In the final, a SQL Dashboard was created, with reproducible notebooks as output, in which plastic packaging weight can be observed according to various filters such as a month, year, and item, showing forecasts for the coming months, showing the top 10 items based on most plastic packaging is used for the current month and next months.

## Outcomes

As a primary deliverable of the project, I constructed a SQL Dashboard that allows the user to see the sold plastic packaging quantity in tones, give estimation till the end of business year for the run-out, and show top items, which past / predicted values are significantly high and should be optimized.

The dashboard presents to the users with the three data cards, general information about the weight of plastic packaging on a monthly and yearly basis. These are total weight of plastic packaging sold for the current year and previous year, the total amount of plastic weight saved during the year, the monthly current plastic packaging weight used according to the selected month and the predicted value for that month.

Secondly, it shows the 10 products with the highest plastic packaging weight on a monthly basis, as well as their estimated value for that month. These are the products that should be optimized in terms of packaging changing.

The monthly ton value of the total weight of the plastic packaging for the current month can be compared with the previous year, and a comment can be made about the monthly values for

this year, and it can also be compared with the forecasted values for the future. At this point, the dashboard presents a line graph showing these three values together.

Another output is that after selecting the product or more than one product with the help of the widget, the dashboard presents the data of this product and last year according to the months. It is possible to see how the sales of the product change according to the months and compare it with the previous year. The last output shows the amount of plastic packaging of the selected product over time and the forecasted data for the future.

## Learning Experience

- As I developed the project using python, my python coding skills have improved. In this process, I gained experience by improving my knowledge about pyspark, pandas, statmodels, and sklearn libraries.
- As a project output, I prepared SQL Dashboard in Databricks platform for the first time, which was a new experience for me. It was a great opportunity for me to work on the Databricks platform.
- In terms of coding, I learned how to design the coding logic to achieve the results you planned in this process.
- I had a little knowledge about time series, from Data Analysis courses at CEU. Thanks to this capstone project, I learned the structure and analysis of time series. I gained information about which tools to use and which methods to use when analyzing time series.
- I learned how to communicate with stakeholders during the project process and how to give them a progress report. In this process, I had many difficulties with the data, especially when I was faced with the data I was going to work on, and we met from time to time to discuss potential outputs and potential solutions.