MS-E2177
Seminar on Case Studies in Operations Research

Data-Driven Optimization of Used Car Inventory

Interim Report

Team:
Juuso Varho (Project Manager)
Tuomas Mäkelä
Patience Anipa
Linh Duong
Juuso Seuri
Client:
Kesko

June 5, 2025

1 Introduction

This interim report provides an update on the progress of K-Auto's used car inventory optimization project. The project focuses on building a data-driven model that recommends the ideal inventory composition for a rolling three-month planning horizon. The model aims to enhance decision-making by prioritizing vehicles based on their potential to maximize return on capital employed (ROCE) while offering justifications for stock inclusion or exclusion.

The report outlines any changes to the project's objectives and scope based on new insights. It also gives details about the current project status, covering completed, ongoing, and upcoming tasks. Additionally, updates to the risk management plan and project schedule are provided to reflect adjustments made during implementation.

2 Revised project plan

The core objectives of the project remain unchanged, with a continued focus on developing a data-driven inventory optimization model to improve Kesko's used car trade by improving profitability and operational efficiency. The model is designed to support purchasing managers in optimizing inventory composition while ensuring alignment with market demand and financial goals.

While ROCE is still part of the plan, it has not yet been implemented. One potential approach involves estimating ROCE by analyzing price changes of categorized car "elements" between purchased and sold datasets, which will be explored further.

Additionally, constraints related to optimization have been identified. The model requires a limited number of elements or categorizations to maintain computational feasibility. Furthermore, scenario modeling relies on clustering techniques to define a manageable number of nodes, and the optimization process operates within a relatively short three-month planning horizon. These constraints introduce trade-offs between model complexity, accuracy, and computational efficiency, which will need to be carefully balanced in the upcoming phases of development.

3 Project status

3.1 Completed tasks

After the course initiation meeting, our group had a meeting with the client to better understand the project objective and the client's needs. After this, a literature review was conducted to find out and learn about inventory optimization methods, and research on used car sales. Additionally, we performed explanatory data analysis on our large data sets, which include what type of cars are currently on sale, what cars have been purchased, and what cars have been previously sold, with the intent of understanding what data is available for our task.

The client's needs were to enhance the decision-making of K-Auto's purchasing managers with data-driven insights on which vehicles should be acquired. Additionally, the model-based suggestions need to have a clear rationale, such as historical data and expected sales performance. This excluded any type of black box models from being used. Based on this requirement, we decided to analyze and predict car sales based on a time-series model, SARIMA. With the SAMIRA model, we can capture and forecast different sales patterns and trends based on seasonality, auto regression, and the moving average.

Car sales data is complex and consists of many categorical and numerical variables, such as the make, model, engine size, color, body type, fuel type, etc. However, this degree of specificity was not required; rather, we wanted to find larger sales patterns, such as the sale prediction of diesel SUV cars. Therefore, each car sale data needed to be categorized into these larger "elements". In order to categorize the sales data, we identified the most important features in the dataset. This was done by calculating feature importance with a Random Forest model, which evaluates how much each feature contributes to the prediction of the target variable. The target variable was chosen to be the asking price of the car divided by the number of days in inventory. With this target selection, we were able to estimate which cars had a relatively high price with regard to the time spent in inventory, where a higher relative value is more desirable. The six most important features were: make, mileage, year, max power, body type, and fuel type. The chosen features were then categorized into 4-6 categories, where an element would be some type of combination of these feature categories, such as {Audi, mileage: 100000-200000 km, year: 2010-2015, 200-250 hp, SUV, diesel}. This set of features contains categorical variables that have more than 6 categories, such as the car make. We kept the 5 best-selling car makers as their own categories and then added an 'Other' category to account for other makers. This categorization can be changed in the future to be clustered, for example, by country or continent of origin, or brand reputation, such as "luxury". The SARIMAX models were then fitted to the elements for demand forecasting.

3.2 Ongoing tasks

We are currently formulating and implementing the optimization problem. Based on the problem, we chose to model the used car sale inventory planning as a multistage stochastic optimization problem. Multistage meaning that an agent makes a sequence of decisions over time, and stochastic, meaning the agent is making decisions in the presence of uncertainty. This closely resembles our problem setup, where the used car dealerships need to make purchase decisions based on uncertain demand. Additionally, the model was designed to be used once a month with a three-month planning horizon. Therefore the months can be modeled as stages, and after each month, we know the number of realized sales.

The model is implemented with Julia's SDDP solver. In order to limit the model complexity, we test methods of limiting the number of states in each stage. Each state represents the realized sales number of each element. However, having the combination of all possible realistic sales numbers would create an extremely large number of possible states. Therefore, we are exploring different methods to limit the number of states while keeping the model realistic. We are currently implementing this with element demand categorization and clustering methods.

3.3 Future tasks

After we have implemented the model, we will assess the effectiveness of the model through testing and validation. After this, we have planned a meeting with the client to present our solution and to ask for feedback on our model and implementation. Based on the model performance and the client feedback, we will seek to improve the model. One area of feedback is the required level of specificity in the car sales predictions, meaning how many features and categories to include in our "element" division. During model improvement, by dividing the workload, we will start to write the final report early.

4 Updated risk management plan

The risk management plan remains largely unchanged, as we believe all previously identified risks are still relevant, even though most have not materialized so far. However, the likelihood of the risk of model not generalizing has been increased to medium. This is caused by the computational complexity of our chosen model approach, as it can only accommodate a limited number of features and categorizations. Our model development is still ongoing and this risk has not yet been realized.

Risk	Likelihood	Effect	Impact	Mitigation
Too ambitious targets	Low	Multiple solutions built, but none meet the requirements	High	Identify which of the problems are the most important and focus on those
Schedule risks	Low	We run out of time and objectives are not met	High	Set clear milestones and schedule "hackathon-style" working sessions for the group
Insufficient understanding of the requirements	Medium	Too much time used implementing complex solutions that do not create value for the client	Medium	Discuss requirements with the client throughout the course
Free riding	Low	Some group members do significantly less work than others	Medium	Schedule regular meetings and agree collaboratively on a fair workload distribution
Model not generalizable	**Medium**	Model built does not create value in real life	Medium	Analyze the data to understand its limitations and conduct rigorous validation

Table 1: Risks related to the project. Updated likelihood written in bold.

5 Updated schedule

Figure 1 shows the timeline for the execution of the project. The key date for the project, which was the crunch weekend scheduled for the 22^{nd} of March was executed successfully and we progressed with the project well enough as planned.

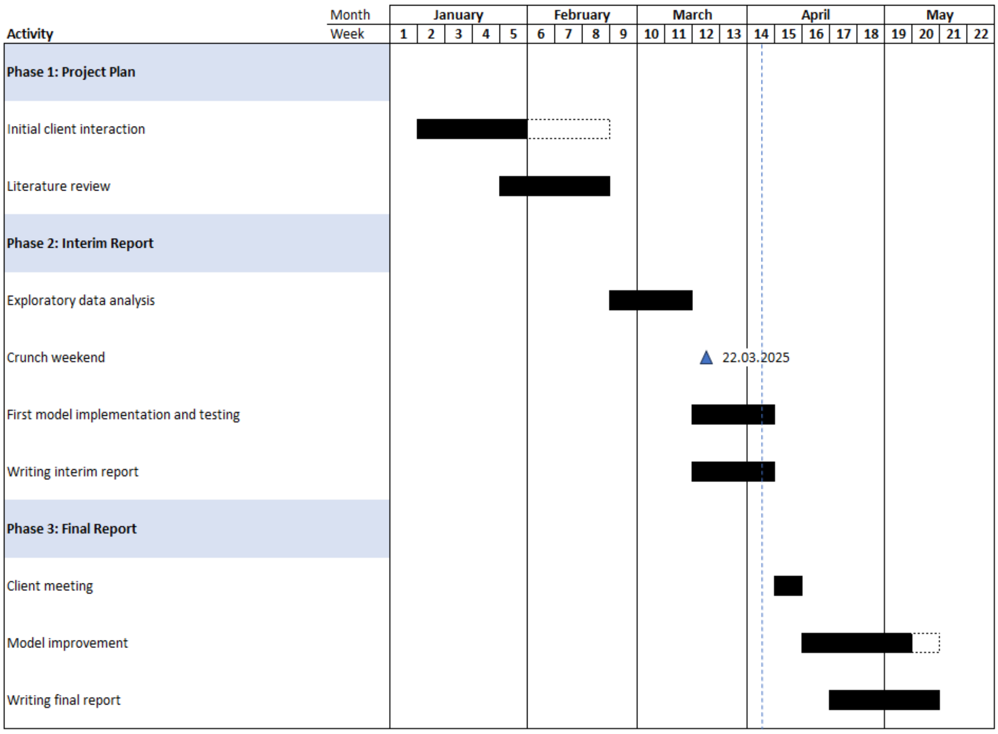

Figure 1: Updated schedule