Saint Joy A. Mandalinao
BSDS 2
February 06, 2025

Homework 1: Introduction to Forecasting

1. For the case studies in Section 1.5, list the possible predictor variables that might be useful, assuming that the relevant data are available.

Case 1

A large car fleet company asked us to help them forecast vehicle resale values. They purchase new vehicles, lease them out for three years, and then sell them. Better forecasts of vehicle sales values would mean better control of profits; understanding what affects resale values may allow leasing and sales policies to be developed in order to maximize profits. At the time, the resale values were being forecast by a group of specialists. Unfortunately, they saw any statistical model as a threat to their jobs, and were uncooperative in providing information. Nevertheless, the company provided a large amount of data on previous vehicles and their eventual resale values.

Possible Predictor Variables:

Vehicle Characteristics:

Model: Some models and makes will retain value more than others.

Year of Production: Newer vehicles will command higher resale prices.

Mileage: Miles on the vehicle will affect depreciation and resale value.

Engine Size: Smaller and larger engines will have different resale values based on demand.

Transmission Type: Automatic vs. manual transmission will have different resale values.

Vehicle Condition: Body, interior, and mechanical condition will affect price.

Color: Some colors will be more in demand in the used vehicle market.

Fuel Type: Gasoline, electric, hybrid, or diesel fuel type will affect resale value based on demand.

Market Factors:

Market Trends: Economy's state, demand for a certain type of vehicle, and general state of the market.

Fuel Prices: Higher fuel prices will affect demand for certain types of vehicles (e.g., SUVs vs. small cars).

Vehicle Popularity: Demand for a certain type of vehicle (e.g., SUVs, electric cars) by consumers.

Supply and Demand: General state of the market's saturation with specific models or types of vehicles.

Leasing Details:

Lease Term: Duration of lease (e.g., 3 years) would affect condition and rate of depreciation of the vehicle.

Mileage Limit: Most leases have a mileage limit, which would affect resale value.

Residual Value in Lease: Price agreed to in the lease, which will affect final resale value.

External Factors:

Economic Conditions: GDP growth rate, unemployment rate, or other economic conditions that can affect consumers' ability to purchase used vehicles.

Seasonality: Time of year (e.g., cars will sell higher during summer or holidays).

Government Policies: Tax benefits for electric vehicles or changes in rules that will affect sales of cars.

Vehicle Use History:

Accident History: Cars that have been in an accident will have lower resale value.

Service History: Routine maintenance and repair records will increase resale value.

Case 2 In this project, we needed to develop a model for forecasting weekly air passenger traffic on major domestic routes for one of Australia’s leading airlines. The company required forecasts of passenger numbers for each major domestic route and for each class of passenger (economy class, business class and first class). The company provided weekly traffic data from the previous six years. Air passenger numbers are affected by school holidays, major sporting events, advertising campaigns, competition behaviour, etc. School holidays often do not coincide in different Australian cities, and sporting events sometimes move from one city to another. During the period of the historical data, there was a major pilots' strike during which there was no traffic for several months. A new cut-price airline also launched and folded. Towards the end of the historical data, the airline had trialled a redistribution of some economy class seats to business class, and some business class seats to first class. After several months, however, the seat classifications reverted to the original distribution.

For the case of forecasting weekly air passenger traffic on major domestic routes, the following predictor variables might be useful, assuming the relevant data are available:

Time and Date Variables:

Week of Year: Useful in capturing seasonality patterns (e.g., holiday peaks, school holidays).

Day of the Week: Passenger traffic may vary between weekdays and weekends.

Month or Quarter: Useful in capturing monthly or quarterly patterns that may be demand-related.

Year: Useful in capturing long-term movements or trends in the market.

Holidays: School holidays, public holidays, and long weekends that may impact travel patterns.

Passenger Class-specific Factors:

Seat Redistribution Data: Seat redistribution from economy to business and first class, or vice versa, that may impact passenger numbers, especially in the pre- and post-change months.

Class-specific Pricing: Variations in ticket prices or discount offers for each class (economy, business, first class) that may impact the split of passengers by classes.

Market and Competitor Behavior:

Competition Activity: Combinational activity of competing airlines (e.g., the cut-price airline) that may impact passenger numbers.

Pricing Strategies: Promotional pricing, fare increases or reductions, or discounts that may lead to demand peaks, especially in specific classes.

Seat Availability: Seat availability in different classes, especially if there are seat reassignments or capacity changes.

Events and External Factors:

Sporting Events: Large sporting events in specific cities that may impact air travel demand.

Festivals and Conventions: Large events such as cultural festivals, conferences, or conventions that may drive demand to travel to specific cities.

Weather Events: Major weather events (e.g., floods, bushfires) or other disruptions that can affect air traffic on specific routes.

Political and Social Events: Elections, strikes, or other events that may impact flight demand.

Operational Issues:

Pilots' Strike or Labor Disruptions: The impact of the pilots' strike or other labor disruptions on passenger traffic.

Flight Cancellations or Delays: Cancellations or delays affect passenger demand in subsequent weeks, especially when customers reschedule.

Economic Indicators:

GDP Growth or Recession: Economic growth can increase disposable incomes, leading to more travel, while recessions can decrease demand.

Unemployment Rates: Lower unemployment can increase travelers, while higher unemployment can reduce discretionary travel.

Airline-specific Data:

Flight Frequency: Frequency of flights per week on every route might have a bearing on passenger numbers.

Seat Capacity: Seat availability or aircraft substitution on specific routes might influence the number of passengers.

Flight Occupancy Rates: Average occupancy rate of flights can forecast passenger demand patterns.

Social and Cultural Factors:

School Term and School Holidays: Differing school calendars in cities might bring about variations in travel demand, especially for family and leisure travel.

Public Awareness Campaigns or Advertising: Advertising campaigns or promotions to specific passenger segments might have an influence on demand.

Historical Data and Trends:

Previous Passenger Traffic Data: Historical passenger volumes, especially seasonality trends from previous years.

Trend Analysis: Long-term patterns in passenger volumes, especially post-strike, after seat reclassifications, or after entry of competitors.

These variables would help the airline develop a comprehensive forecasting model to predict weekly air passenger traffic on major domestic routes and manage seat allocation across different classes.

2. For case 1 in Section 1.5, describe the five steps of forecasting in the context of this project.

Case 1: Section 1.5 A large car fleet company asked us to help them forecast vehicle resale values. They purchase new vehicles, lease them out for three years, and then sell them. Better forecasts of vehicle sales values would mean better control of profits; understanding what affects resale values may allow leasing and sales policies to be developed in order to maximize profits. At the time, the resale values were being forecast by a group of specialists. Unfortunately, they saw any statistical model as a threat to their jobs, and were uncooperative in providing information. Nevertheless, the company provided a large amount of data on previous vehicles and their eventual resale values.

Five Steps of Forecasting in the Context of Vehicle Resale Values (Case 1, Section 1.5)

Step 1: Problem Definition

The company needs to forecast the resale values of vehicles after a three-year lease period by creating a data-driven forecasting model based on extensive historical data of previous vehicles and their eventual resale values. This task is challenging due to the lack of cooperation from internal specialists, who view statistical models as a threat, potentially leading to gaps in qualitative insights that explain certain data variations, and the inherent complexity of the factors affecting resale values—such as economic conditions, brand perception, mileage, and overall vehicle condition. To address these challenges, the approach begins with a clear problem definition that outlines the forecasting objective and identifies key stakeholders who will use the forecasts to refine leasing and sales strategies.

To ensure a smooth transition, the model should be positioned as a decision-support tool that enhances expert judgment rather than replacing it. Additionally, integrating the model into existing workflows with user-friendly dashboards and providing specialists with an option to adjust predictions based on real-world conditions will encourage adoption and collaboration.

Resale values can also be significantly impacted by variables such as consumer preferences, fuel prices, and economic trends. A model should include data validation, preprocessing, expert consultations, and historical trends and macroeconomic indicators to ensure accuracy. Comparing expected and actual resale values can be facilitated by frequent updates and a feedback loop. Integrating an advancement or continuously improving forecasting system can make an impact on improving decision-making, which can also lower financial risks, and maintain the company's competitiveness.

Step 2: Gathering Information

Gathering Information, the focus is on assembling a comprehensive and detailed dataset that will form the backbone of the forecasting model. Two types of data are required in order to develop a high-precision automotive resale value prediction model: statistical data (quantitative data) and expert opinion (qualitative data). They can be integrated, so there will be a balanced approach to forecast resale values and to measure measurable and experience factors. Information gathering, the focus is on collecting a comprehensive and thorough dataset that will serve as the foundation of the forecasting model. This involves the collection of quantitative information from historical information, including vehicle-specific information (e.g., make, model, year, mileage, and purchase price), lease duration, maintenance record, and ultimate resale values, which hold the key information on depreciation patterns and resale performance over time. It is also useful to collect contextual information relating to economic indicators (e.g., interest rates, inflation, and regional economic cycles), market conditions, and brand reputation, since these are bound to have a significant influence on resale values. Given the recalcitrant nature of internal specialists is taken as a given, an effort should be made to elicit any available qualitative observations from documentation, such as maintenance records, customer complaints, or earlier internal reports, and supplement these with external market surveys and industry analyses. This multi-dimensional data gathering ensures that both quantitative patterns and qualitative nuances are addressed, providing firmer ground for subsequent exploratory analysis and model building.

Step 3: Preliminary (Exploratory) Analysis

Preliminary (Exploratory) Analysis, consists of the primary task of visually and statistically exploring gathered data to identify underlying trends, patterns, and anomalies. It is started through data visualization—plotting resale values against time, vehicle characteristics, or economic conditions—to observe if there are any apparent trends, seasonal trends, or cyclical trends. By creating graphs such as line charts, scatter plots, or box plots, we can identify whether there is a consistent pattern of depreciation, seasonal highs or lows, and outliers that significantly deviate from the mean. To properly analyze the data, various visualization techniques will be used to identify trends and patterns in car resale values. Time series plots will be used to track how resale values have changed over time, and scatter plots can identify correlations between the most significant variables such as mileage and resale value. Box plots will also identify outliers, such as some car models retaining value significantly better or worse than expected. Identifying patterns in the data is crucial, such as trend analysis to track whether resale values are increasing or decreasing, seasonality analysis to identify if times of the year influence prices, and anomaly detection to account for unusual circumstances such as economic recessions. Correlation analysis will identify the strongest correlations between variables, such as the negative correlation between mileage and resale value or the positive effect of fuel efficiency on resale prices. Data quality, data cleaning and processing will involve dealing with missing data using imputation techniques and correcting or removing outliers that can mislead the model's accuracy.

Besides visualization, the step involves conducting a comprehensive statistical analysis to identify the correlation between variables. Correlation analysis can be employed to identify which vehicle characteristics (such as mileage, make, or model) influence resale values the most, while summary statistics (mean, median, standard deviation) provide an overview of the data distribution. Furthermore, the identification of outliers or data points with abnormal behavior may result in a further investigation such as re-checking qualitative sources or collecting more context to explain these anomalies. This comprehensive exploratory analysis forms a basis for choosing the appropriate models in later steps by bringing into focus the prevailing factors that influence vehicle resale values.

Step 4: Choosing and Fitting Models

Selecting the appropriate forecasting model depends on the available data and the problem's complexity. In Step 4, Choosing and Fitting Models, focus shifts to selecting the most appropriate statistical or machine learning model based on the findings gathered from exploratory analysis. Several factors influence this selection, including whether historical data are available, the quality of such data, the nature of interdependence between vehicle attribute and resale values, and what use forecasts are to be made of in the future. For instance, if there are primarily linear relationships between variables, then multiple linear regression will suffice, but if there are non-linear interactions, or complex relationships, present in the data, then more sophisticated methods like decision trees, random forests, or gradient boosting algorithm might be more appropriate and Time Series Model If historical resale values show trends or seasonal patterns. The model also needs to be considered in terms of interpretability versus predictive power, particularly if domain experts are likely to be suspicious of statistical models. A combination of different models can be used to improve prediction accuracy and once a model is chosen, it needs to be trained on historical data, validated using test data, and optimized by tuning parameters to improve accuracy.

Step 5: Using and Evaluating the Forecasting Model

Once the model is implemented, the assessment starts: It entails confirming accuracy and reliability by comparing predicted resale values with actual outcomes of resold cars. This assessment is useful for spotting any bias or inconsistency in the model, so adjustments may be made, including tuning parameters, adding extra variables, or even going backward in data-processing stages. The changes ensure responsiveness to new data trends and maintain consistency, creating an ongoing feedback loop which provides trustworthy insights for decision-making at the end.