A picture containing icon

Description automatically generated

| *Student Full Name* | **Yuri Braga** |
| --- | --- |
| *Student Number* | **sba24328** |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 2 |
| *Assessment Due Date* | *15th December 2024 (23:59 IST)* |
| *Date of Submission* | Friday, 29th November 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

### **1. Strategic Overview of the Business Problem**

The used car market is expected to reach $460 billion by 2029, registering a CAGR of over 10% over the forecast period. And with demand expected to continue growing throughout 2025, accompanied by rising new car prices, still a reflection of the COVID pandemic that has stalled the auto industry, there is a certain inclination among consumers to buy used cars, strengthening the entire sector. The growth of digital platforms for buying and selling used cars has further strengthened consumer confidence in buying used vehicles.

While the industry keeps growing, more dealerships are joining Saas solution platform to process online orders and transactions. It is noticed that there is a need for a tool that can be used to price a used car.

For instance when a customer offers their own car in exchange as part of the payment, it is difficult for a seller to understand whether the price that was sent or asked for the used car is compatible with a list price for that model.

This capstone project aims to deliver an accurate price prediction tool in the used car market. Once implemented, this tool would be available specially for dealerships to handle vehicle pricing. The tool can offer several benefits as such:

**Environment and Sustainability**

Accurately priced used cars can help with environmental sustainability as a longer vehicle lifespan reduces the need for new car production which involves higher emissions and uses of natural resources.

**Customer Satisfaction**

When a customer submits a price for their used car or is looking to purchase one, it is important that the price is fair to both parties (buyers and sellers). The price assigned needs to be in line with the market, ensuring that the consumer feels confident in their decision.

**Market Efficiency**

Appropriate pricing helps the market remain efficient and stable. It helps to maintain stable levels without excessive prices causing a distortion of the market with unfair buying and selling conditions.

**Increase profitability**

Since accurate price prediction allows for more effective management of inventory control. By applying real market values ​​to vehicles, dealers can improve their inventory levels and work on turnover strategies, thus increasing overall profitability.

### **2. Project Plan**

This project aims to cover the following deliverables:

* A deployed model that predicts used vehicle price
* A simple user interface to interact with the model

This capstone project will use the Agile philosophy and Scrum methodology.

Developed in the early 1990s, Scrum is an Agile framework that helps to generate value through its adaptive solutions for complex problems.(Schwaber and Sutherland, 2020).

Scrum Framework

When applying Scrum on a project, the load of work is divided into Sprints, which are fixed-duration iterations and it typically lasts two weeks. Scrum involves different roles and process:

* A product owner requires the work for a complex problem creating a product backlog.
* The scrum team turns a selected part of this into an increment of value during the sprint.
* Stakeholders along with the scrum team will review the results and if necessary adjust the next sprint.
* The process should repeat until the goal is accomplished.

Implementation

Timeline and Sprints

This capstone project aims to have two-week Sprints. On each sprint here is the list of deliverables:

Sprints

| 1 | 2 | 3 | 4 | 5 | 6 |
| --- | --- | --- | --- | --- | --- |
| Data Acquisition and EDA | Data cleaning and preprocessing | Exploratory Data Analysis and Initial Modeling | Advanced Modeling and Initial Results | Model Refinement and Validation | User interface, Documentation and Deployment |

Sprint 1

Goal: Acquire necessary vehicle datasets and conduct initial exploratory data analysis to get information about the data and its patterns.

Sprint 2

Goal: Data cleaning and processing. Important to remove missing values or incorrect data. Perform the necessary data transformations to prepare data for model training.

Sprint 3

Goal: EDA to refine hypotheses and initial predictive models built.

Sprint 4

Goal: Develop more complex models. Implement analytics techniques to improve prediction accuracy.

Sprint 5

Goal: Refine models by conducting validation to ensure model reliability.

Sprint 6

Goal: Create a detailed documentation with the EDA, report and findings. Develop a simple user interface to interact with the model. Deploy the tool.

**Resources**

**This capstone will use Python for data cleaning and model implementation.**

**It will be delivered as a Jupyter Notebook. State libraries.**

### **Business understanding**

**This capstone project would directly impact decision-making and improve profitability.**

**Key impacts this tool would bring:**

**Accurate pricing**

**Undervaluing or overprice a car can disrupt the transaction and bring user dissatisfaction. This tool will enhance pricing accuracy which is essential to improve customer trust and increase transaction volumes.**

**Operational Efficiency**

**Once a trade-in is received, Dealerships can assess those and set fair sales prices quickly reducing the time vehicles spend in inventory reducing holding costs.**

**Higher Profitability**

**When the price asked for a vehicle is closely aligned with the market conditions and car valuations, dealerships will optimize their profit margins.**

**Liability**

**On digital marketplaces, providing a pricing tool that offers AI can help to stardize car prices which leads to market transparency. This leads to greater trust among end users, attracting even more traffic to the platform.**

**Single point data for negotiation**

**Both ends, buyers and sellers can use data from the same tool to inform their negotiations, which leads to a more balanced and fairer transaction based on current market tendencies.**

#### **Stakeholders**

**The primary beneficiaries of the used car price prediction tool include:**

**The price prediction tool will benefit the following stakeholders:**

**Digital selling platforms**

**Digital marketplaces that enable users to buy or sell their used cars can integrate this service to deliver instant listings pricing.**

**Dealerships selling used cars**

**Being the primary users of the tool, Dealerships often need to provide instant evaluation over listings on a daily basis. This process might happen multiple times along the day and require precision with an accurate result. This predicting tool can complement or replace any already in place pricing methodology with the goal to maximize profitability and improve efficiency.**

**Individual sellers**

**Private users that want to trade-in their used vehicle can use this tool to get instant evaluation before a trade-in is submitted. This brings a fair negotiation with potential buyers with a better understanding of their vehicle value.**

### **3. Business Understanding**

* **Stakeholders: Identify who will benefit from this tool (e.g., individual sellers, used car dealerships, online marketplaces).**
* **Business Impact: Discuss how improving price prediction accuracy can impact decision-making and profitability in the used car market.**

### **4. Data Understanding**

Data Source

The data set used during this study is the “Car details v3.csv” extracted on [car-dataset](https://www.kaggle.com/datasets/nehalbirla/vehicle-dataset-from-cardekho) which is subject to the Open Database License (ODbL). In accordance with such license the users are allowed to distribute, alter and utilize the database but observing the licensing conditions for attribution requirements and akin sharing constraints. This dataset is very relevant for the study as it contains adequate and organized information on used vehicles, which is indispensable in the creation of a potent price estimation tool.

Licensing details:

The dataset is made available under the Open Database License (ODbL) v1.0 by the Open Knowledge Foundation.

This license makes provisions for the right to:

* Apply the database for any purposes, whether commercial or noncommercial.
* Change and develop new works from this database.
* Replicate or pass the database or its modified works on to third parties under the same licensing regime.

The conditions include:

* Attribution: credit must be given to the original author of the database for each use or derivative work.
* Share Alike: ODbL licensing that is attributed on any changes made or any derivative database created must be sustained on that same ODbL licensing.

Compliance Plan:

Concerning the licensing restrictions, this project will: Adhere to the licensing restrictions, and apply appropriate strategies. Proper acknowledgment of origin of sources will be made by ensuring that relevant details such as names of datasets and links to their licensing terms are provided in the documentation of the project.

The dataset contains the history of used car sales collected from various online selling websites. These sites gather and store lots of data pertaining to sales and listings making it a comprehensive data set of used cars market in all the territories over a wide time frame and includes many details.

Features Description

The data set contains a number of essential characteristics which are of great assistance in predicting the resale price of second-hand cars:

**Name:** The name and model of the car (for example, Maruti Swift Dzire’s model is VDI).

**Year:** This indicates the age of the car in years during which it was manufactured.

**Selling\_Price:** This is the price or the listing under which the car can be offered for sale.

**Km\_Driven:** This is the total kilometers driven by the car and is useful to assess wear and tear.

**Fuel:** This includes the type of fuel the car uses such as diesel, petrol, or LPG.

**Seller\_Type:** Here, the type of seller may either be an individual or a dealer.

**Transmission:** This refers to the type of transmission available such as manual or automatic.

**Owner:** Ownership status such as first-owner or second-owner.

**Mileage:** basically, consumption of fuel expressed in km per liter for petrol or km per kg for Highlander.

**Engine:** This is engine size capacity in CC.

**Max\_Power:** This refers to maximum power out from the engine in BHP.

**Torque:** It is the torque produced by the engine usually represented in Nm or Kgm.

**Seats:** It refers to the number of seats the vehicle has.

* **Initial Observations: Share any initial insights from a cursory look at the data.**

### **5. Data Preparation**

* **Cleaning Steps: Explain how you handled missing data, outliers, and errors.**
* **Feature Engineering: Describe any new features you created and why.**
* **Data Transformation: Document any transformations or scaling applied to the dataset.**

### **6. Exploratory Data Analysis**

* **Descriptive Statistics: Provide statistics that summarize the central tendency, dispersion, and shape of the dataset's distribution.**
* **Visualizations: Include charts and graphs to illustrate relationships between features and the target variable.**
* **Preliminary Findings: Highlight any significant trends or patterns you observed.**

### **7. Findings & Recommendations**

* **Model Performance: Discuss how the models performed, including metrics like RMSE or R².**
* **Insights: Share insights on what drives car prices based on your model’s findings.**
* **Recommendations: Offer suggestions for potential buyers, sellers, or platforms based on your findings.**
* **Future Work: Propose further areas of research or additional features to improve the model.**

### **8. Version Control and Repository**

* **GitHub Usage: Explain how you used GitHub for version control.**
* **Repository Link: Include a link to the GitHub repository containing your Jupyter Notebook and report.**

### **9. References**

* **Citation: Ensure all your sources are properly referenced using the Harvard referencing style. Include any datasets, articles, or books you utilized.**

### **Word Count and Submission**

* **Adherence to Guidelines: Keep the report within the 2,000-word limit, focusing on clarity and conciseness.**
* **Professional Presentation: Ensure the report is well-organized, properly formatted, and free of grammatical errors.**

### **Tools and Tips for Execution**

* **Jupyter Notebook: Use this for all your coding, analysis, and visualization tasks.**
* **Libraries: Utilize Pandas for data manipulation, Matplotlib/Seaborn for data visualization, and scikit-learn for modeling.**
* **Iterative Approach: Continuously refine your analysis based on the findings and feedback.**

**This structured approach will help you create a comprehensive and coherent report that meets the requirements of your assignment while providing valuable insights into the used car market.**

**Assessment Task:**

This assessment aims to evaluate your ability to apply project management methodology to develop and execute a capstone project. You will select a dataset, conduct exploratory data analysis, pre-process the data, implement at least one machine learning algorithm, and present your findings effectively through a comprehensive report. The capstone project will be based on a dataset of your choice from any domain, such as finance, marketing, or any other.

You will

submit a comprehensive report detailing the following;

* Strategic overview of the business problem Project plan
* Business understanding
* Data understanding
* Data preparation
* Machine learning implementation Findings
* Conclusions
* Any future recommendations

The report should be presented in a clear and concise manner, and it should demonstrate your ability to use a project management methodology. The project management methodology should enable you to prioritize tasks and monitor the progress of the capstone project. Additionally, the report should provide a background to the business problem and its importance from a strategic viewpoint, an overview of the project's timeline, milestones achieved, and any challenges faced during the implementation phase. It should highlight the key insights gained from analyzing the data and present any significant trends or patterns observed. Moreover, the report should address any limitations or constraints encountered during the project and propose potential solutions for future improvements.

This is a two-semester module, and the capstone project will continue into semester two. Students are advised to review and adhere to the submission requirements documented after the assessment task.

Further details of the assessment:

* a) **Continue to use the GitHub repo provided in CA 1, the Jupyter Notebook and report**    
  **Word document must be put into a GitHub repo for version control. The GitHub repo link will be added at the end of the report. There should be another 5 to 10 commits throughout the time worked on CA2.**
* **b) Exploratory data analysis of your dataset. Use descriptive statistics.**
* c) **Use at least one machine learning algorithm.**
* d) **Support your analysis with references and properly reference ALL sources that you**    
  **have used. WARNING – If you do not support your work, you will not receive a high mark!**
* e) WORD COUNT: 2,000 words. If your report is too short or long, you may *lose up to 10% of marks*!

**Assessment Requirements**

All assessment submissions must meet the following minimum requirements:

* ● Include a professional report paper in Word format ONLY of about 2,000 words.
* ● Code must be submitted as a Jupyter Notebook artefact.
* ● ZIP or RAR files will not be accepted. Files must be submitted separately.
* ● Be submitted by the deadline date specified or be subject to late submission penalties.
* ● Be submitted via Moodle upload.
* ● Use Harvard Referencing when citing third party material.
* ● Be the student’s own work.
* ● Include the CCT assessment cover page.  
     
  **Learning Outcomes:**    
  This assessment addresses the following module learning outcomes for this module:  
   1. Critically evaluate the relationship between information technology infrastructure and organisational competitive advantage.  
   2. Critically analyse and select open source and proprietary software with a view to developing IT  
   solutions for business and business-related IT problems.  
   3. Utilise tools of strategic business analysis to evaluate the current macro and micro business environment with a view to formulating future action plans.  
   4. Research emerging technologies and critically evaluate their impact on business and business information systems in general.  
   5. Understand the relationship between data gathering/utilisation and business intelligence and its impact on industry policy.

### **~~1. Strategic Overview of the Business Problem~~**

* **~~Context~~**~~: Start by describing the used car market and why accurate price prediction is critical.~~
* **~~Problem Statement~~**~~: Define the specific problem your project addresses—predicting the resale value of used cars.~~
* **~~Importance~~**~~: Explain why this is important for buyers, sellers, and businesses within the automotive industry.~~

### **~~2. Project Plan~~**

* **~~Goals~~**~~: List the primary goals of your project.~~
* **~~Timeline~~**~~: Outline the timeline of your project with key milestones and deadlines.~~
* **~~Resources~~**~~: Describe the tools and technologies (e.g., Python, Jupyter Notebook, specific libraries) you will use.~~

### **3. Business Understanding**

* **Stakeholders**: Identify who will benefit from this tool (e.g., individual sellers, used car dealerships, online marketplaces).
* **Business Impact**: Discuss how improving price prediction accuracy can impact decision-making and profitability in the used car market.

### **4. Data Understanding**

* **Data Source**: Describe where your data comes from and why it is suitable for this project.
* **Features Description**: Provide a detailed description of the dataset features (e.g., make, model, mileage, year, etc.).
* **Initial Observations**: Share any initial insights from a cursory look at the data.

### **5. Data Preparation**

* **Cleaning Steps**: Explain how you handled missing data, outliers, and errors.
* **Feature Engineering**: Describe any new features you created and why.
* **Data Transformation**: Document any transformations or scaling applied to the dataset.

### **6. Exploratory Data Analysis**

* **Descriptive Statistics**: Provide statistics that summarize the central tendency, dispersion, and shape of the dataset's distribution.
* **Visualizations**: Include charts and graphs to illustrate relationships between features and the target variable.
* **Preliminary Findings**: Highlight any significant trends or patterns you observed.

### **7. Findings & Recommendations**

* **Model Performance**: Discuss how the models performed, including metrics like RMSE or R².
* **Insights**: Share insights on what drives car prices based on your model’s findings.
* **Recommendations**: Offer suggestions for potential buyers, sellers, or platforms based on your findings.
* **Future Work**: Propose further areas of research or additional features to improve the model.

### **8. Version Control and Repository**

* **GitHub Usage**: Explain how you used GitHub for version control.
* **Repository Link**: Include a link to the GitHub repository containing your Jupyter Notebook and report.

### **9. References**

* **Citation**: Ensure all your sources are properly referenced using the Harvard referencing style. Include any datasets, articles, or books you utilized.

### **Word Count and Submission**

* **Adherence to Guidelines**: Keep the report within the 2,000-word limit, focusing on clarity and conciseness.
* **Professional Presentation**: Ensure the report is well-organized, properly formatted, and free of grammatical errors.

### **Tools and Tips for Execution**

* **Jupyter Notebook**: Use this for all your coding, analysis, and visualization tasks.
* **Libraries**: Utilize Pandas for data manipulation, Matplotlib/Seaborn for data visualization, and scikit-learn for modeling.
* **Iterative Approach**: Continuously refine your analysis based on the findings and feedback.