Weird Motors Pricing Calculator

C964 - Computer Science Capstone

Jaxon Weis

June 2022

Submission 1

TABLE OF CONTENTS

A1 - Letter of Transmittal

Jaxon Weis
Chief Technology Officer
Weis Industries
101 Technology Circle
Forney, TX 75126

Dear Mr. Weird and board members,

The automotive market has had a turbulent time this past year. The availability of cars and car parts has had a great impact on the sales of many automotive companies. Due to this turbulence, automotive companies must use creative marketing, innovative design, and competitive pricing to stay relevant. Your company is known for designing high-quality vehicles with a range of features. However, over the past year, Weird Motors has seen declining sales due, in part, to the lack of competitive pricing. One solution to aid your design and development team is a calculator that would allow teams to estimate the market value of a vehicle based on the vehicle's features.

This calculator would use machine-learning capabilities to analyze data and calculate pricing based on factors such as: vehicle type/body style; features, cost of competitors, etc. Using this tool, Weird Motors could distribute data to internal departments such as design and development, marketing, manufacturing, etc. to allow these departments to make decisions on how to increase profits based on consumer stats and competitor models.

The funding required to implement this pricing calculator tool will be approximately $45,000. This funding is required for development costs, data costs associated with using a third-party researcher, and man-power related costs. It will include the first 18 months of development, quality assurance, and platform implementation of the proposed project.

I believe that my team has the ability to customize an incomparable machine learning product that will not only increase your profit margins, but also set your company apart from competitors.

I would love to schedule a meeting to discuss my proposal and your specific needs. Please do not hesitate to contact me with any questions.

Sincerely,
Jaxon Weis
Chief Technology Officer

A2 - Project Proposal

Problem Summary

In today's market, an important factor in business is staying competitive. It isn't good enough anymore to have cars that people won't buy due to non-competitive pricing. Weird Motors needs a solution that will allow the company to stay competitive and relevant in an ever-changing market with ever-changing prices. For this reason, the company would benefit from a calculator that will generate market-fair pricing for new vehicles based on the features they have. This will ensure that any car produced will be priced competitively

Benefits to the Customer/ Decision-Making Process

This calculator will input available features and estimate a price based on other car prices in the market. With this estimate, the cars produced are priced competitively so they can be sold off of the lot sooner. This estimate is even more beneficial when used during the design phase. This calculator will support the decision-making process by providing a tool to estimate a car's value based on the features we design, which will enable the company to tweak the design as needed. The calculator will allow the design team to more effectively increase profits by analyzing the cost of manufacturing, shipping, marketing, etc. versus the cost consumers will pay based on market value, availability, and competitor pricing.

Outline of the Data Product

The base of this project is the dataset. It is necessary to have solid data in order to train the model. When using data in this way, the first step is to check to make sure it's healthy, and to visualize the data to see trends. The data is then randomly split between training data and testing data. Training data is used to create a model and testing data is used to check the model. Once a successful model is generated, it can be used to predict pricing for any car based on the features.

Description of the Data

The dataset used to create the calculator was sources from Kaggle.com. The data used has 205 entries of cars, features and prices. Each entry has these columns:

- Car_ID object
- CarName object
- Symboling int64
- Fueltype object
- Aspiration object
- Doornumber object
- Carbody object
- Drivewheel object
- Enginelocation object
- Wheelbase float64
- Carlength float64
- Carwidth float64
- Carheight float64
- Curbweight int64

- Enginetype object
- Cylindernumber object
- Enginesize int64
- Fuelsystem object
- Boreratio float64
- Stroke float64
- Compressionratio float64
- Horsepower int64
- Peakrpm int64
- Citympg int64
- Highwaympg int64
- Price float64

The data is packaged in a CSV file which will upload into the program easily. This data is not perfect. One of the issues is some misspellings of the car names. In most cases that would have to be corrected, but since car names are not needed for the calculator, the column was dropped altogether. Car_ID was also dropped because it is not useful for our calculator. The rest of the columns, besides price, are used as the independent variable features. Price will be the dependent variable.

Objectives and Hypothesis

The objective of this proposal is to provide a calculator tool that will accurately predict the market value of a car based on competitor pricing and features of the car. This will allow Weird Motors to create a competitive pricing plan and increase market share and profits.
The hypothesis for this proposal is: If the data of features and pricing of competitive cars is used to train a Linear Regression machine learning model, then a calculator tool can accurately estimate the market value of a car based on features listed in the data.

Product Methodology

The methodology approach used in this project is SCRUM. This was chosen due to the limited time of this project and the organization of a list of tasks to be completed and checked off when done. This also helps since the limited personnel on this project. A sprint to work on the back end of the project can be done separately from the front end. This also works with Jupyter notebook since the notebook is an iterative process that can be refined as time goes on. Due to SCRUM being related to agile, SCRUM benefits from interacting with the customer more often to create the customer's vision. SCRUM is broken down as follows

1. Product backlog - This is the list of tasks of the entire project sorted by priority.
2. Task Backlog - This is the list of tasks currently in progress and should be completed soon.
3. Implementation - This is the process of the tasks being completed and implemented into the project.
4. Review - Daily all of the team members will get together to explain what they were able to get finished, what they are working on and if they have any roadblocks.

Funding Requirements

The product development will require initial funding of $45,000. The initial funding will cover the full development of the product, which includes the software licenses, equipment, and

the time used by the data analytics team, software design team, and quality assurance/platform implementation specialists. If additional requirements that are out of scope for the product are added during the development cycle, additional funding will be required. If additional requests are made, the additional funding will be calculated and the requests will be analyzed for effectiveness and benefits to the overall product before the agreement will be amended.

Impact on Stakeholders

The stakeholders will now be able to price their up and coming vehicles without risking pricing themselves out of the market. This will allow for Weird Motors to move products quicker. If this tool is used in the design process for the customers' new cars, the calculator would be able to check profit values and then tweak the design to make the design more profitable. Overall the impact should increase profit and decrease overhead costs, while providing more transparency to stakeholders.

Ethical & Legal Considerations / Precautions for working with sensitive data

The dataset used in this product is publicly available. There is no reason to be concerned about the data used for this product. The data does not involve any personal information or trade secrets that would cross an ethical or legal line. The development of this product has considered and adhered to a variety of regulations including the following acts: HIPAA, FERPA, PCI DSS, or the Privacy Act of 1974.

In order to maintain ethical, legal, and moral compliance, all parties included in the design, development, and usage of this product will be required to complete a 1-hour course on information privacy as well as review and sign a non-disclosure agreement with Weis Industries confirming that they understand and will uphold all policies and procedures set in place by Weis Industries to maintain the integrity of the data and any proprietary information.

Personal Relevant Expertise

The lead developer in this project is well known for his management and leadership skills. He also has a strong background in computer science with an abundance of technical experience. He will be leading a small team of college interns to turn the project goals into reality. The team members are well experienced in python and algorithm design.Utilizing the internship program will allow this project a more creative and innovative design process as these students will be familiar with up-and-coming technology resources and have a fresh take on how established companies can be more competitive to a younger market. This project will benefit both the lead developer and the team of interns because it will require learning of a new client and the business needs in the automotive industry. All team members should expect to see professional and personal growth and development throughout the lifecycle of this project.

B. Executive Summary

Opportunity

Weird Motors is an established automotive company known for designing high-quality vehicles with a range of features. Over the past year, Weird Motors has seen declining sales due, in part, to the lack of competitive pricing. One solution to aid this company is a calculator tool that would allow them to estimate the market value of a vehicle based on the vehicle's features. By analyzing the data generated by the pricing calculator, Weird Motors would maximize productivity in their business decision-making and offer competitively priced vehicles with an emphasis on price by feature which will increase market share and overall profits while decreasing overhead costs.

Description of Customer & Needs

The pricing calculator tool will be used by the design, development, marketing, and manufacturing departments of Weird Motors. It will allow the Weird Motos team the ability to analyze and utilize data in a new way by directly comparing vehicle costs by feature of direct competitors. Interested parties will be trained on how to generate reports and calculate hypothetical scenarios to inform their decisions according to market trends and business needs. The pricing calculator also provides data visualization so that the customer can quickly and easily identify trends.
Stakeholders will be crucial participants in the development of this product. The pricing calculator development team will meet with stakeholders at the beginning of each project milestone to provide an update as well as gain feedback on the project scope as it relates to the needs of the business.

Existing Gaps in Data Products

One of the issues the dataset had when first working on it was some of the car manufacturer names were misspelled. The original plan was to fix the misspellings in the dataset. Later in the design process we discovered that the manufacturer name didn’t help us complete our goal so the column was eventually dropped. The other hurdle we had was some of the features were listed as strings. This was corrected by using dummy variables so they would be able to be used in the machine learning model.

Data to Support the Data Product Lifecycle

To ensure that the model accurately calculates market value new datasets must be formed to keep the most up to date information. This will also help when new cars are hitting the market to include those pricing models. To combat these issues a new data set will be made once every other month to keep the data fresh. It would also be recommended to do an off cycle dataset update if there are significant changes to the market.

Methodology for Design and Development

This project will be maintained under the design methodology called SCRUM. The purpose of using SCRUM compared to Agile was due to the time factor. Both SCRUM and Agile are both designed to work well for an iterative design process. The reason SCRUM was chosen was due to the sprints. I believe these sprints will deliver this project faster compared to agile. With SCRUM there is a backlog of tasks to be completed. Whenever anyone has completed their own task they can easily identify a new task to complete as well.

Deliverables

The product deliverable will be the access to the calculator on a webpage gui. Since the entire program is not being handed over Weis Industries will be responsible for maintaining the system and updating the dataset. Due to this subscription model, nothing that is created in-house is lost. This model could even be used by multiple companies if multiple contracts are established.

Plan for Implementation & Anticipated Outcomes

The plan for this project is to list all individual tasks needed to get the project complete. Having the SCRUM master being able to issue tasks and remove obstacles so that the project can be completed as quickly as possible. The use of daily SCRUM meetings will be beneficial for the success of the project.

At the end of this project the outcome should be a fully functional car price calculator embedded into a webpage that can be accessed by our customers. The deliverable of a webpage is better than an application due to being in control of the software produced. It also limits the risk of information being leaked by reverse engineering applications.

Methods for Validating the Developed Data Product

One of the better characteristics of the SCRUM method is that we can stay in constant communication with the customer to make sure the customer is satisfied with the project through each milestone of the project. The project will also be tested through the various stages of the project to reduce the amount of bugs in the project.

The Weird motors price calculator has to be checked to make sure the model works appropriately. The prices that the calculator generates has to be compared to market rates. If the model is not predicting price as expected the dataset might need to be expanded or modified depending on how the model is affected.

Programming Environments & Costs

Programming Environment

Weird Motors Pricing calculator is being built on python 3 inside of Jupyter notebooks. The project uses multiple different libraries including Sklearn's Machine Learning library. This project will also be using github for a repository. To display the jupyter notebook Binder will be used. The site hosting and domain registration will cost $5,000. This will allow us to keep the tool online for our clients.

Human Resources cost

The total man-hours used by the development team, quality assurance team, and training resources team is estimated at 280 hours ($28,000) over the 11 weeks of development and implementation with an estimated 120 hours ($12,000) allotted to the quality monitoring period of 6 months. This totals $40,000 in human resource costs.

Projected Timelines & Milestones

The timeline and milestones are reflected in the table below. This project will take approximately 11 weeks to complete with a 6 month quality monitoring period after full implementation to ensure that the product is running efficiently and the client's needs are met. The quality monitoring period will consist of fortnightly meetings between the QA team and stakeholders followed by meetings between the QA team and the development team if necessary. All tasks should be completed in order unless otherwise noted in the dependencies column.

PHASE & TASK NUMBER	EVENT	START DATE	END DATE	DURATION (IN HOURS)	DEPENDENCIES	RESOURCE ASSIGNED
P1 T1	Project Requirement Meeting	07/01/2022	07/01/2022	4	N/A	Development Team, Quality Assurance Team, Stakeholders
P1 T2	Project Kickoff Meeting	07/05/2022	07/05/2022	2	P1 T1	Development Team, Quality Assurance Team, Stakeholders
P1 T3	Product Planning	07/06/2022	07/10/2022	20	P1 T3	Development Team
P2 T1	Product Design	07/11/2022	07/24/2022	40	All of P1	Development Team
P2 T2	Data Cleaning	07/22/2022	07/25/2022	20	All of P1	Development Team
P2 T3	Data Model Trials	07/26/2022	08/05/2022	20	P2 T1-T2	Development Team, Quality Assurance Team

P2 T4	User Interface Trials	08/06/2022	08/10/2022	20	P2 T1-T2	Development Team, Quality Assurance Team
P2 T5	Debugging / utilizing feedback in prototype design	08/10/2022	08/19/2022	20	All previous tasks	Development Team, Quality Assurance Team
P3 T1	Deliver Prototype	08/22/2022	08/22/2022	2	All of P2	Development Team, Quality Assurance Team, Stakeholders
P3 T2	Usability testing	08/22/2022	09/05/2022	30	P3T1	Development Team, Quality Assurance Team
P3 T3	Feedback meeting	09/05/2022	09/05/2022	3	P3 T1-T2	Development Team, Quality Assurance Team, Stakeholders
P4 T1	Incorporate feedback into product	09/05/2022	09/10/2022	20	All of P3	Development Team
P4 T2	Develop Customer Training for user interface	09/05/2022	09/12/2022	30	All of P3	Development Team, Training Resources Team
P4 T3	Quality Testing	09/10/2022	09/14/2022	20	P4 T1	Quality Assurance Team
P5 T1	Deliver Product to Client	09/15/2022	09/15/2022	10	All of P4	Development Team, Quality Assurance Team, Stakeholders
P5 T2	Acceptance Testing and Debugging	09/15/2022	09/18/2022	15	P5 T1	Development Team, Quality Assurance Team, Stakeholders
P6	Full implementation	09/20/2022	09/20/2022	4	All of P5	Quality Assurance Team, Stakeholders
P6 ONGOING	Quality Monitoring	09/20/2022	03/20/2023	120	All of P5	Quality Assurance Team, Stakeholders
			TOTAL	400		

C. Design & Develop

Descriptive & NonDescriptive Method

The non-Descriptive data analysis used in this project are the graphs displayed on the Jupyter paper. In the paper I have shown both scatter plots and a heatmap. This will help Weis Industries understand the data being used. Hopefully this data will give early warning signs showing how well the model will perform.

On the Descriptive side of things the calculator will be using Linear Regression to estimate the price of cars. Once the user inputs the features of a car onto the calculator the model will be able to predict the price of that data.

Datasets

The dataset (Kumar, 2019) used for this project was publicly sourced from Kaggle.com. Kaggle's user got the data from archive.ics.uci.edu. The link to the dataset is provided: https://www.kaggle.com/datasets/hellbuoy/car-price-prediction

Decision-support functionality

This project is designed for its decision-support functionality. The biggest asset being the market value of the car. With that information, a company can better decide on how to establish a competitive pricing model. This gives a good starting point in the seller's process. This project can also benefit in the design phase of a car. If the car being designed does not meet the profit goals the designed can be changed to meet those goals

Featuring, Parsing, Cleaning, & Wrangling Datasets

Most of the original dataset (Kumar, 2019) was kept. Two items were parsed from the dataset: the Car name and the Car ID. Neither one of those attributes helped to estimate price, so those items were dropped due to lack of value to the end product.

Methods & Algorithms for Data Exploration & Preparation

In the project the dataset attributes were split between number and string variables. From there, scatter plots of the numbered variables were generated. The scatterplots showed the trend of different variables as they relate to price. The string variables were put into strip plots to show the density of a feature at different price points.

Data Visualization Functionalities

Every one of the attributes listed in the dataset was entered into a graph to see the impact on price. The pricing calculator includes a variety of data visualization functions by utilizing scatterplots, strip plots, and a heatmap. All of these graphs describe the relationship between that attribute and the price of the car. One of the more interesting correlations I see is a negative relationship between horsepower and price.

Interactive Queries

The included gui allows a user to input all of the features of a car to estimate the price. Features can be interchanged to generate varying price points for hypothetical scenarios.This allows the user to make the informed decision by giving the user the market value of the car.

Machine-Learning Methods

The machine learning method used is linear regression. This was used due to the ability to train the model on a large number set. The linear regression was chosen for its ease of use and simplicity. This method provides clear results in a user-friendly format which will benefit both the development team and the client.

Evaluating Accuracy

Accuracy is determined by the linear regression's own scoring tool. This tool generated a score based on its ability to estimate the testing data. This tool will score the model somewhere between 0 and 1. The score is based on the R^2 number that signifies how well an equation fits trended data.

Security Features

Since all of the assets in this project are digital, there is a threat of information being leaked. Digital security will be of the utmost importance. Not only for our team members but for our clients as well. Username and encrypted high security passwords are necessary for the continuation of this project.In addition to technological security measures, ethical/legal measures have also been taken such as providing data privacy training and requiring an NDA.

Monitoring & Maintaining

Continual monitoring of our application and machine learning models will be prioritized to keep the operation running efficiently. The Models score will be continuously checked to make sure it is still working correctly.

User Dashboard

The user GUI is the reason clients will use the tool. The professionalism and quality of the dashboard will maintain our clients satisfaction to keep them coming back. The dashboard includes many of the visualizations of the entire dataset along with the query option of estimating the price of a car by its features.

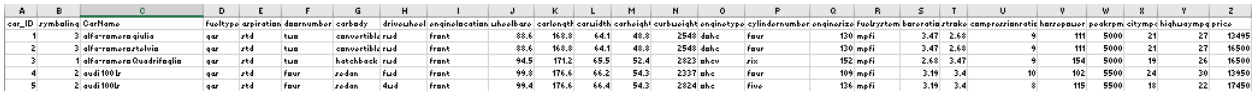
D. Documentation

Business Vision

This project seeks to achieve the goal of creating a competitive pricing model for new cars based on the features of the car in order to increase revenue and decrease overhead costs. In order to meet this goal, the company needs a product that will support business decision making. The pricing calculator tool is a value-added interface that allows the client to create a fair and competitive pricing model by using data from the current automotive market. This calculator will save time and resources by providing an easy-to-use interface that allows users to input available features and estimate a price based on other car prices in the market. With this estimate, features can be added or subtracted from newly manufactured cars and all cars produced are priced competitively from the beginning so they can be sold more quickly, which will provide a quicker return on investment and decrease storage costs.By analyzing the data generated by the pricing calculator, Weird Motors will increase market share and overall profits while decreasing overhead costs.

Raw & Cleaned Data Sets

The raw dataset was discovered publicly on Kaggle.com. The data was stored in a CSV, comma separated values, file seen below.



A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
car_ID	symboling	CarName	fueltype	aspiration	doornumber	carbody	drivewheel	enginelocation	wheelbase	carlength	carwidth	carheight	curbweight	enginetype	cylindernumber	enginesize	fuelsystem	boreratio	stroke	compressionratio	horsepower	peakrpm	citympg	highwaympg	price
1	3	alfa-romero giulia	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	13495
2	3	alfa-romero stelvio	gas	std	two	convertible	rwd	front	88.6	168.8	64.1	48.8	2548	dohc	four	130	mpfi	3.47	2.68	9	111	5000	21	27	16500
3	1	alfa-romero Quadrifoglio	gas	std	two	hatchback	rwd	front	94.5	171.2	65.5	52.4	2823	ohcv	six	152	mpfi	2.68	3.47	9	154	5000	19	26	16500
4	2	audi 100 ls	gas	std	four	sedan	fwd	front	99.8	176.6	66.2	54.3	2337	ohc	four	109	mpfi	3.19	3.4	10	102	5500	24	30	13950
5	2	audi 100ls	gas	std	four	sedan	4wd	front	99.4	176.6	66.4	54.3	2824	ohc	five	136	mpfi	3.19	3.4	8	115	5500	18	22	17450

To clean the data set we needed to remove unnecessary variables. The Car ID and Car name columns were not necessary for generating the market values so they were dropped. The string variables were incompatible with the machine learning library so they were converted into dummy variables. The cleaned data set shown below.

	symboling	wheelbase	carlength	carwidth	carheight	curbweight	enginesize	boreratio	stroke	compressionratio	...	cylindernumber_three	cylindernu
0	3	88.6	168.8	64.1	48.8	2548	130	3.47	2.68	9.0	...	0	
1	3	88.6	168.8	64.1	48.8	2548	130	3.47	2.68	9.0	...	0	
2	1	94.5	171.2	65.5	52.4	2823	152	2.68	3.47	9.0	...	0	
3	2	99.8	176.6	66.2	54.3	2337	109	3.19	3.40	10.0	...	0	
4	2	99.4	176.6	66.4	54.3	2824	136	3.19	3.40	8.0	...	0	

Code Used

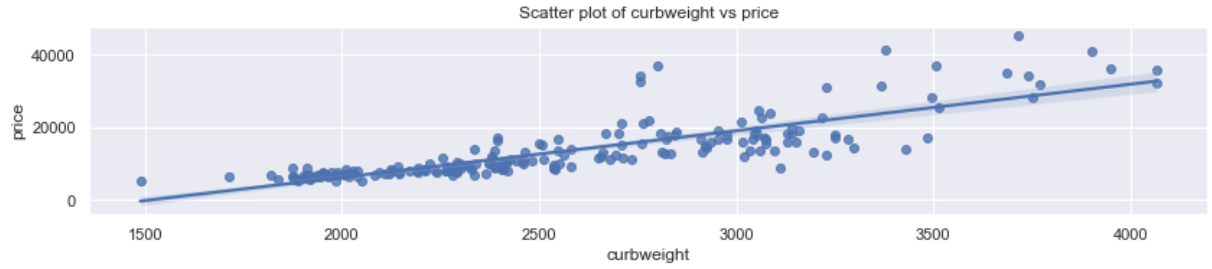
The Code in this project is relatively straight forward, with only a few libraries used. The first being pandas, which is used to create the data frames in which the dataset is stored. One we have the data stored we can use visualizations to explore the data. The project is using seaborn for all of the visualizations. Using the visualizations makes the modifying of the data easier. 2 items were identified to not help with the calculations one was the Car ID the other car name.Those items were removed since they were not needed. After the data frame was inputted into a function from sklearn to randomly separate the training data from the testing data. The training data is then fed into the training model.

Assessment of Hypothesis

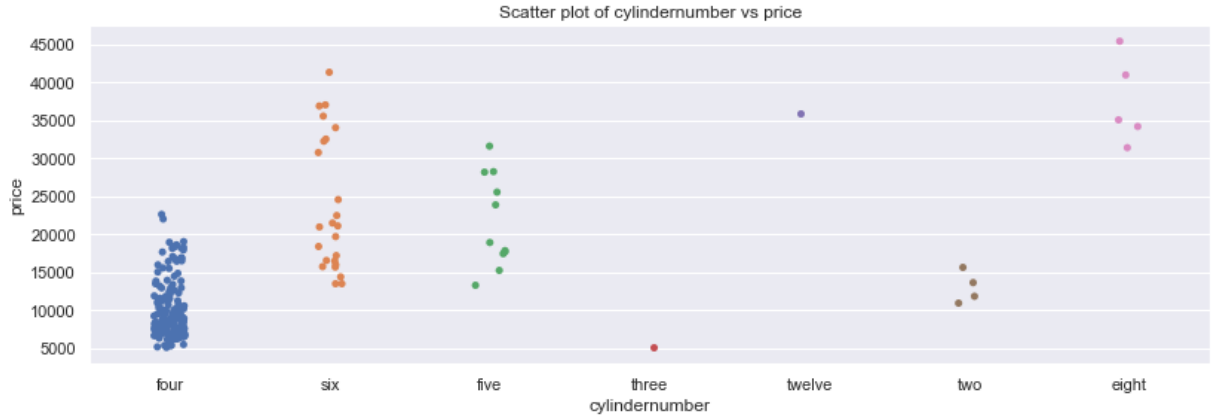
In the project we were successful in creating a machine learning model that is predicting market values of cars based on their features. The score output of the model rests at a 0.93. Which in turn calculates the car prices within a 30% range. Something that we can do to improve that number is increase the learning data and add a column for the class of manufacture for example luxury, mass maker, or sport.

Supporting Visualizations

A strong correlation between curb weight and price.



The correlation between cylinder number and price.



Looking through the Jupyter files of this project every characteristic is put on a graph compared to price. This will give the user and techs ideas of how the model will behave when generating market values. This will be extremely important to the team members working on this project.

Product Accuracy

The accuracy of the model is very important to its success. To keep the model accurate the project will have the dataset continually updated with the latest market trends. The model R^2 score currently sits at 0.93. When the model is applied to the testing data the market value was off at most 30%. That could be improved with a larger dataset.

Results & Revisions

The end of this phase of the project has created a working proof of concept that is able to generate market values with acceptable tolerance. This project will be dramatically improved with a redesigned GUI compared to the one currently. This project can also be improved performance wise by implementing a bigger data set. This model currently uses 205 records. That is only the tip of the iceberg compared to all of the data available.

Source Code & Executable Files

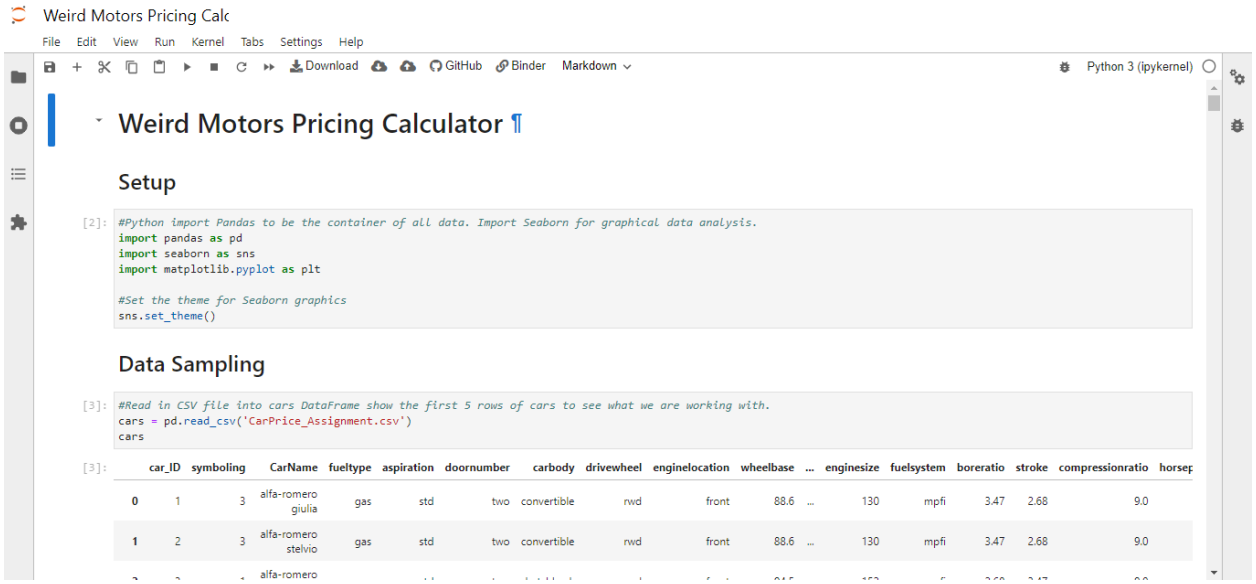
The code used in this project is loaded into a Jupyter file which is stored publicly on github. Link here:
https://github.com/JaxonWeis/C964/blob/main/Weird%20Motors%20Pricing%20Calculator.ipynb

The mybinder link is here
https://mybinder.org/v2/gh/JaxonWeis/C964/main?labpath=Weird%20Motors%20Pricing%20Calculator.ipynb

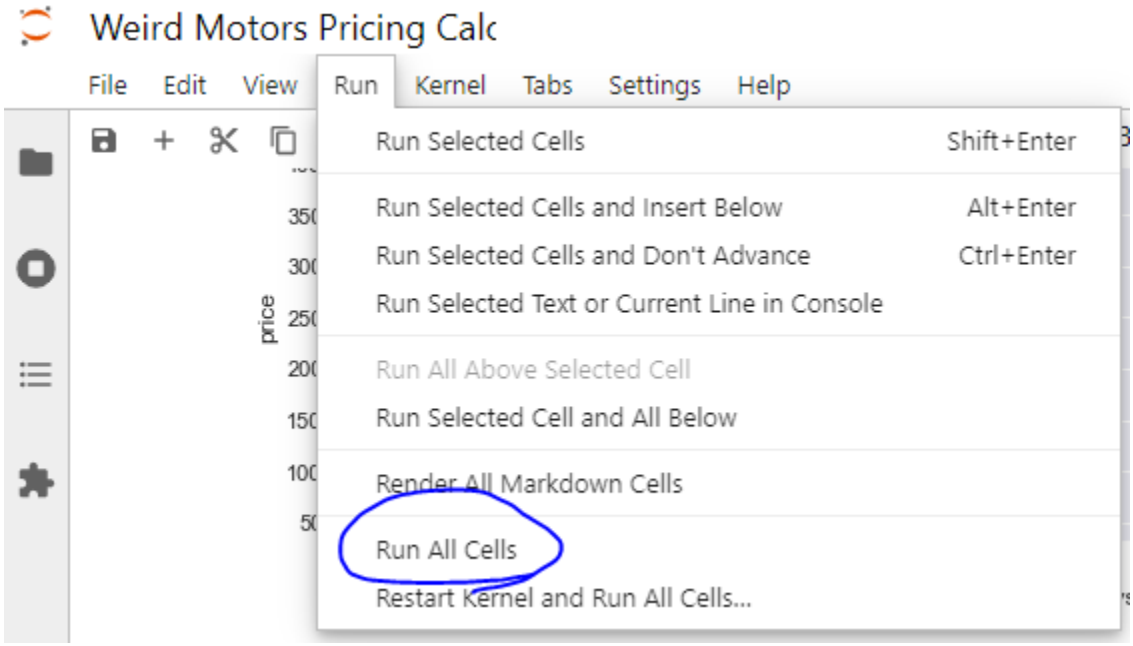
Quick Start Guide

The first step in running the application is this link: https://mybinder.org/v2/gh/JaxonWeis/C964/main?labpath=Weird%20Motors%20Pricing%20Calculator.ipynb

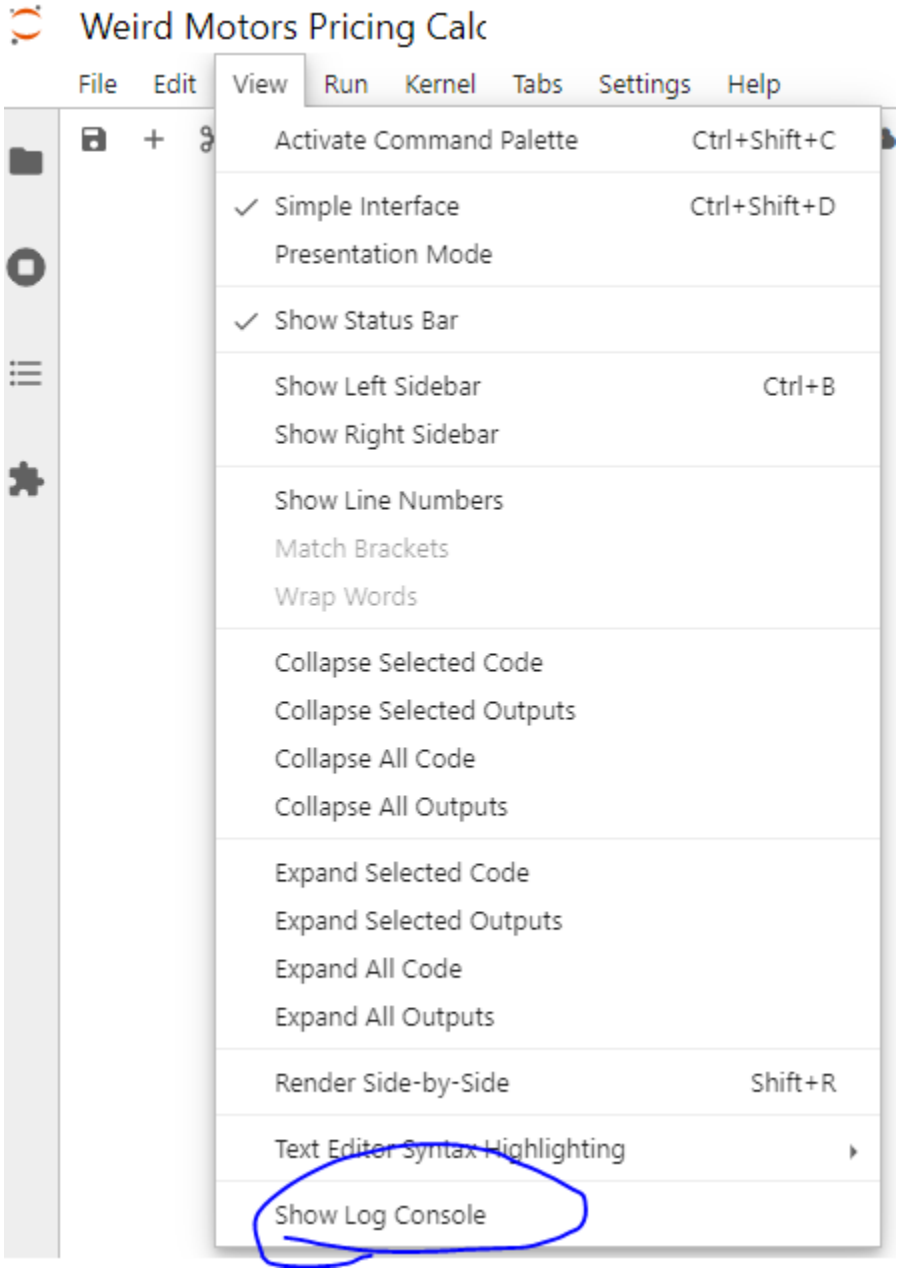
This will bring you to the mybinder site which is setting up the jupyter file. This could take some time approx 5 mins. Once the page loads and you see the Weird Motor Price calculator you can begin.



At the top toolbar you can see the run dropdown option. Click run and then click run all. See image below.



The application will start to run through its pacing. This includes generating graphs and visualizations. When you scroll all the way to the bottom you will be able to see the GUI section of the jupyter file. The last thing we need to do before running the calculator is open the log console. Again at the top toolbar hit view and Show Log Console.



Once that window is open at the bottom we can generate some market rates by entering the characteristics in the GUI section. When finished hit the submit button at the bottom of the GUI page. This will generate a price in the log console. You can keep tweaking different features and generating new market values.

E. Sources

Kumar, M. (2019). *Car Price Prediction Multiple Linear Regression*. Kaggle. Retrieved June 22, 2022, from https://www.kaggle.com/datasets/hellbuoy/car-price-prediction