# 1. General Information

This document describes the specification of our machine learning (algorithm) model and its testing strategy. Keep in mind that the results of the algorithm validation should also serve as input to be referenced as a support for a business plan. Nothing is official yet, until a practical field test is conducted. Under no circumstances use this as the main decision maker for a business. The reason why we are using this business case as a test for implementing machine learning is because we had a brainstorming meeting about how to potentially use the Parc to build new products or services:

A screenshot of a computer

Description automatically generated

What was presented to TILLETT:

A screenshot of a computer

Description automatically generated

A Machine Learning Daily Rate could be a new income source for SMMT:

A screenshot of a search engine

Description automatically generated

# 2. Tillett Case

On November 23rd, during a visit to Tillett with Afam, we had the chance to learn about their business model and future, such as exploring the B to C market. Tillett is a company that specializes in race vehicle seats, with kart seats as their main source of income. The company plans to expand its racing seats to regular customers and sell them on other platforms, or even on their own platform. Cutting off the middleman. According to the CEO's son, they were having trouble comprehending what to do with the data. That’s when I suggested to have a look at our Parc data and possibly use machine learning for Data Mining. See their seats in a regular vehicle below:



**Relevant Machine Learning Tools:**

* Python (Coding)
* SQL (Data Storage)
* Excel (Data observation)
* HTML Browser (Map)
* Task Scheduler (Automation Trigger)
* Miror (Presentation + Planning)

# 3. Development Resources

## This research was intended to address a query raised by the son of Tillett's CEO concerning ways to enhance their sales. Find out where there are potential hot spots in the UK for vehicle owners who own sport car types that may be able to change their vehicle seats.

## 4. Developer Team

| Name | Role |  |
| --- | --- | --- |
| Diego Alves | SDI Product Implementation and Support Officer |  |
| Tim Bruin | SDI Product Manager • SMMT Data Intelligence |  |

## 5. Machine Learning Libraries

| Name |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Webbrowser  Matplotlib  Pandas  Statsmodels  Sqlalchemy |  |  | Numpy  Collections  Ipywidgets  Hdbscan  Folium | Re  KMeans  Make\_blobs  ConvexHull | IO  Plotly  Collections |

## 6. Data

| Count | Description |
| --- | --- |
| 23 columns - 1000 rows | Parc |

**Vehicle Type Sector Make Range Variant Country of Origin Body Style DVLA Body Type Transmission Nominal Engine Capacity Engine CC Fuel Type Aspiration Drive\_Type Power BHP MVRISPostcode Company\_Private Number Previous Keepers Number of Seats Colour Weight Year of 1st Reg Count of Registrations**

CARS C LOWER MEDIUM SEAT LEON LEON CUPRA 290 TSI SPAIN 5 HATCHBACK 5 DOOR HATCHBACK SEMI AUTO 2 1984 PETROL TURBO 4X2 285.6 SL1 C 0 4 BLACK 1356 2020 31

## 7. Development Planning

**Intended Purpose**

Clustering Parc into groups of vehicles that could potentially meet the desire sales group for the sport seats offered by Tillett. Then mapping based on the district post code location.

**How Tillett could use this model to understand the market?**

If a hotspot location has a high number of vehicle groups that could have new sports seats. Tillett could have the potential to use the sales market strategy in these locations to increase sales in the region. Rather than wasting money and resources on other locations where there are no vehicles that can use their seats.

**Software Architecture and Resource Time**

CRISP-DM Model(Cross-industry process for data mining). (2 days)

Extracting the data from SQL Server using Python. (1 hour)

Python will split the data. (1 hour)

Python will apply Linear Regression. (1 hour)

Python will save the data in excel. (1 hour)

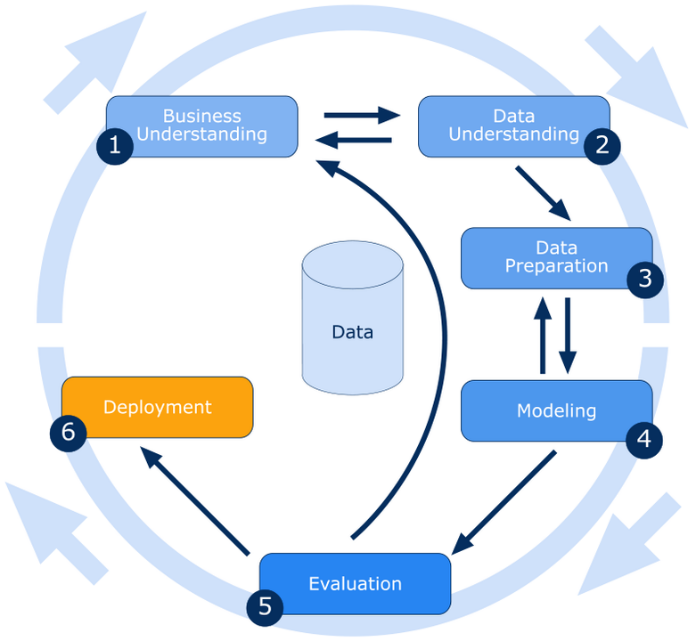
Python will map the data using HTML. (5 hours)

Python will put the data into charts. (2 hours)

Evaluating and making further analysis. (1 day)

**Total: 3 day + 11 hours average**

CRISP-DM Model(Cross-industry process for data mining)



# 8. Algorithm Model Training and Description

Used Linear Regression to find the correlation between columns. The code will use only 1000 records. In this business case: Body Stye, Company/Private filtered by Registrations higher than 10 vehicles from Year of First Registration higher than 1990. Then clustering using K-means and applying to a base map.

# 9. Algorithm Model Evaluation

The model showed some hot spots location across the map. It can be automated buy running every time our new Parc data is uploaded into the server. Chart the registrations.

Map:

A map of the world

Description automatically generated

Linear Regression Table: Ordinary Least Squares (OLS)

A screenshot of a computer

Description automatically generated

Post Code vs Registrations

A graph with lines and dots

Description automatically generated with medium confidence

Variant vs Registrations

A graph with lines and dots

Description automatically generated with medium confidence

Post Code District vs Year of 1st Registration

A graph with colorful lines and numbers

Description automatically generated

# 10. Conclusion

We are currently conducting a test. Exploring other approaches is necessary. The maximum sales point will be achieved with the help of a group of decision makers and a machine algorithm. Making a final decision takes time.