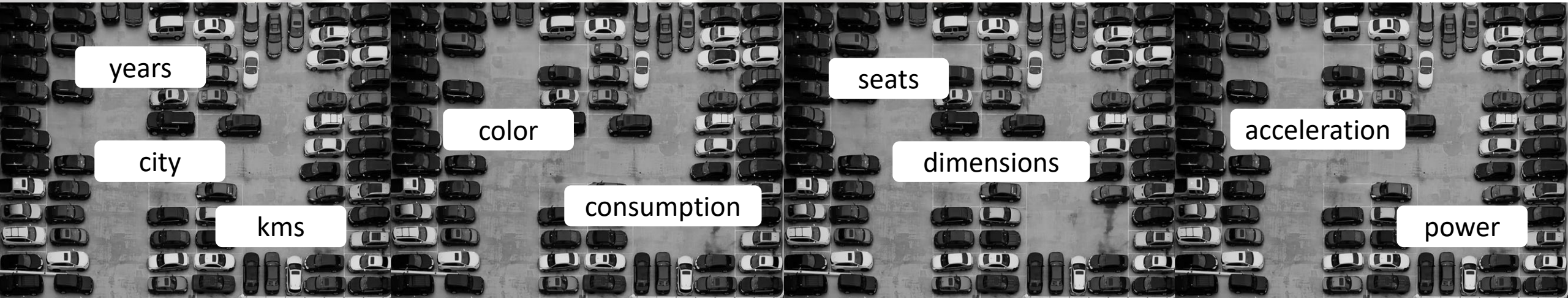
Car Value Appraiser

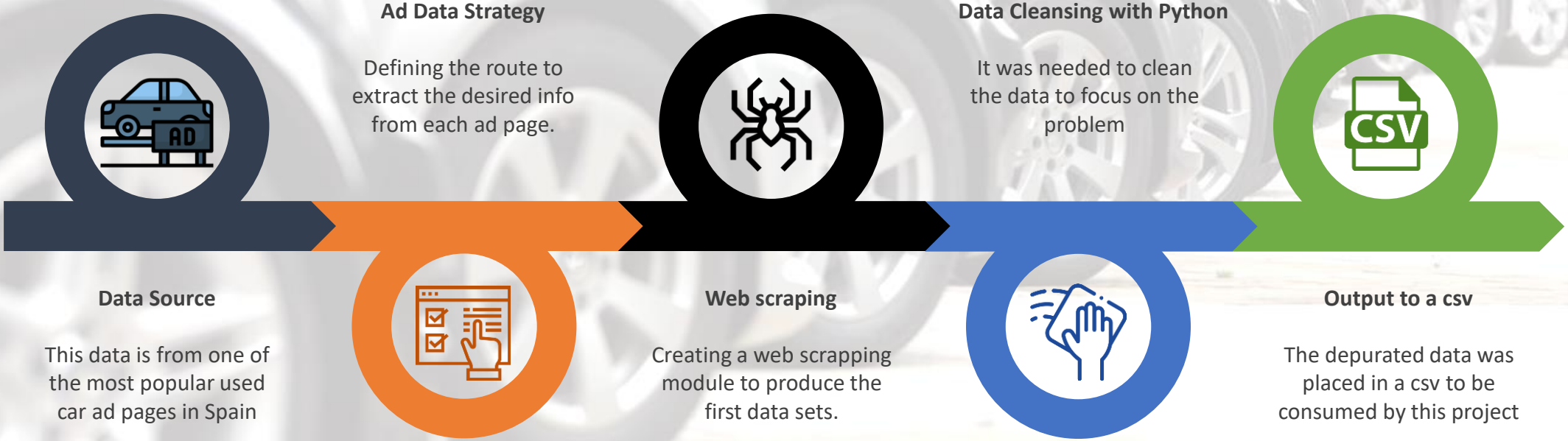
Carlos Espejo Peña

Introduction to the problem

- There is an uncertainty on the price to be paid for a second hand vehicle
- Goal: Create a ML Used Cars Appraiser
- How do the characteristics of the vehicle impact the final price?



Data collection process



Data Source

This data is from one of the most popular used car ad pages in Spain

Ad Data Strategy

Defining the route to extract the desired info from each ad page.

Web scraping

Creating a web scrapping module to produce the first data sets.

Data Cleansing with Python

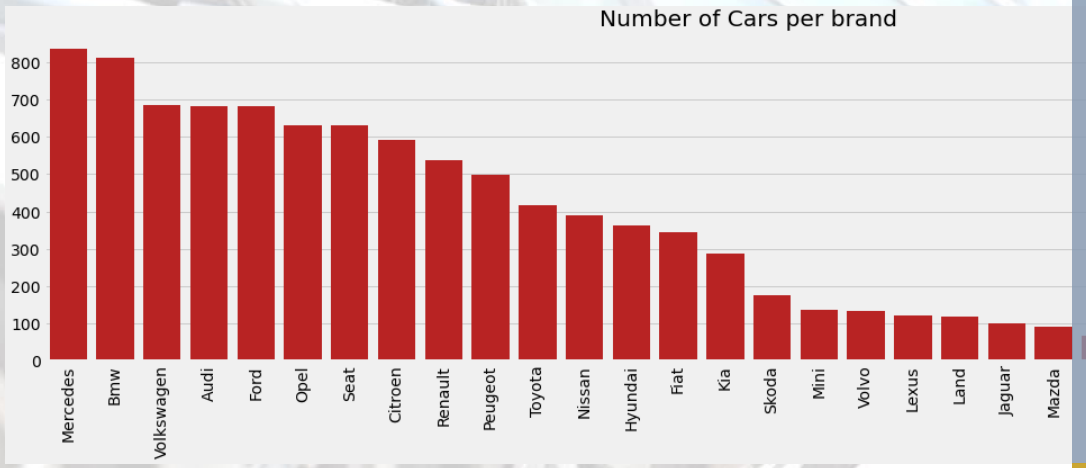
It was needed to clean the data to focus on the problem

Output to a csv

The depurated data was placed in a csv to be consumed by this project

A quick look at the data

- Data from 8 types of cars
- + 20 tech parameters
- Data from renting plans
- Extraction date: 9 Mar 2021
- + 80,000 records



Features (X)

y

	title	year	kms	city	gear	doors	seats	power	color	co2_emiss	brand	model	price
0	Toyota Yaris Hsd 1.5 City	2016	40.000 km	100 cv en Madrid	Automática continua, secuencial	5 puertas	5	100 cv	Blanco	75 gr/m	Toyota	Yaris Hsd 1.5 City	9.500€
1	Mercedes Clase A A 35 Amg 4matic+ 7g-dct	2019	24.000 km	306 cv en Salamanca	Automática secuencial	5 puertas	5	306 cv	Blanco	167 gr/m	Mercedes	Clase A A 35 Amg 4matic+ 7g-dct	43.890€
2	Mercedes Clase A A 250e 8g-dct	2020	1.500 km	218 cv en Tenerife	Automática secuencial	5 puertas	5	218 cv	Gris	32 gr/m	Mercedes	Clase A A 250e 8g-dct	35.900€
3	Volkswagen Polo 1.0 Tsi Sport 70kw	2017	53.000 km	95 cv en Palmas, Las	Manual	5 puertas	5	95 cv	Blanco	106 gr/m	Volkswagen	Polo 1.0 Tsi Sport 70kw	8.990€
4	Dacia Sandero 1.0 Ambiance 55kw	2018	33.600 km	75 cv en Palmas, Las	Manual	5 puertas	5	75 cv	Blanco	117 gr/m	Dacia	Sandero 1.0 Ambiance 55kw	5.990€

Available data

The datasets contain technical specifications from several car brands. Only the most popular ones will be selected

Forecast

The aim of this project will be to build a ML model which could help to predict the price (y) based on the features (X)

Possible Technical Solutions

±1,000 €

Forecast Price Goal

The main objective of this Project is to get to a model that can provide an error deviation < 1k€

EDA with Pandas

- Statistical data analysis.
- Detect the main features that could conform our model.

Machine Learning Algorithms

- Regression: RandomForest, Multivariate Linear Regression
- Neural Networks: MLP, CNN, BNN (Deep Learning: LSTM)
- Classifier: RandomForest, SVM, Decision Tree

Dashboard: Play with data

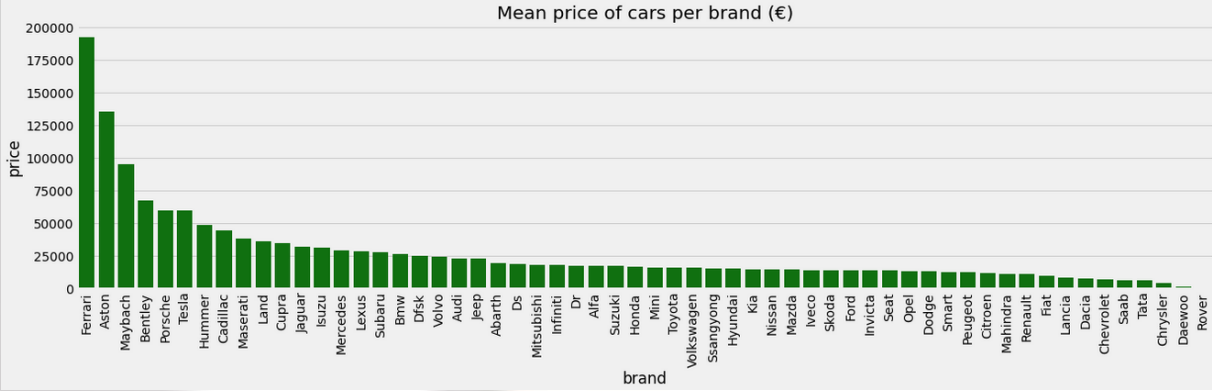
- Set a dashboard with different ML algorithms to be selected
- Create an interactive tool: dash, django, streamlit

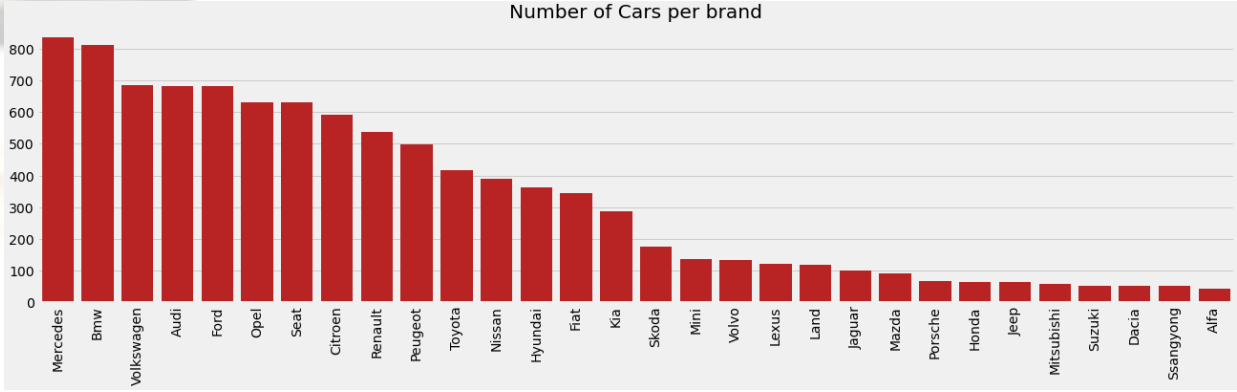
Draw the attention of the audience

- Design a Tableau story. To present all the main insights
- Without a good storytelling, data results could be confusing

Detected Risks

- Differences in prices could alter the direction of the prediction
- Different number of samples for each brand
- Renting could be more convenient for cars of high value





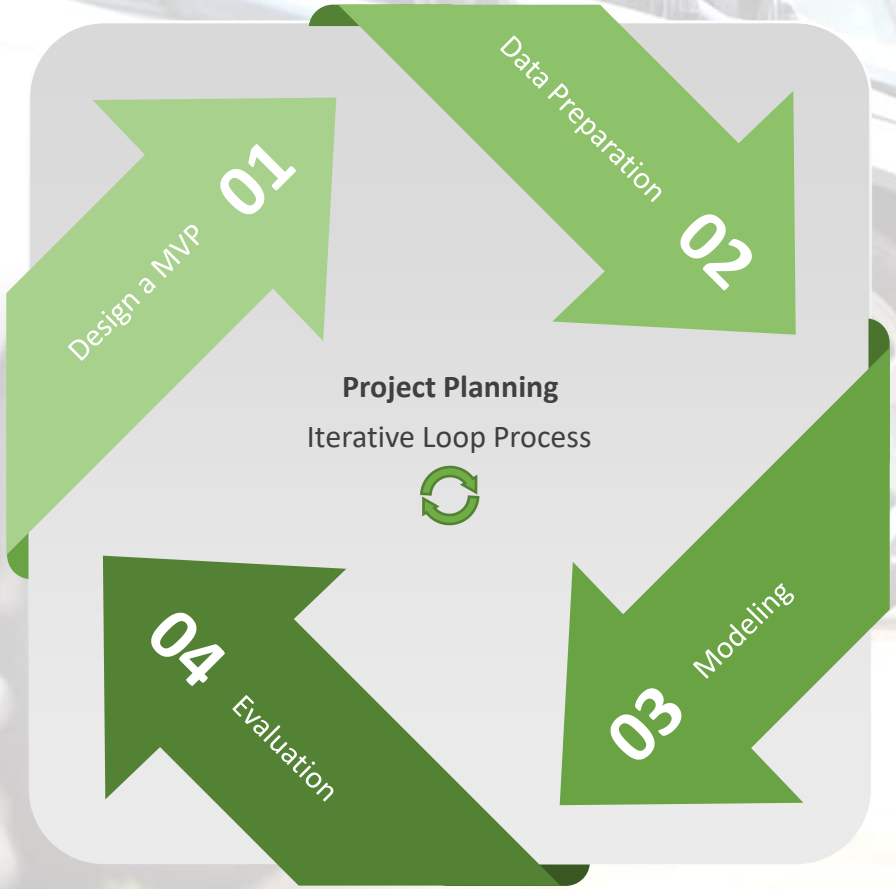
Action Plan

Design a MVP: 15/03/2021

This will consist on building a first quick ML model to predict the data. It delimit the floor of the prediction model that we can obtain

Model Evaluation: 15/05/2021

Design Evaluation functions in python that can provide the same rating to every model. This will avoid inconsistency on the final reported values



Data Preparation: 30/03/2021

Data wrangling process to decide which of the data could be useful to introduce to the model. Statistical analysis will help to see the relation among variables and evaluate their impact to the final forecast

Modeling: 30/04/2021

Investigate about several ML models that could fit to find the solution to this problem. Apply them and see results compared to the actual values

Data Extraction – 10/03

Develop First Model – 31/03

Dashboard Draft – 15/05

Solving Issues & Improve the model – 15/06

Delivery – 22/07

MVP Design – 17/03

Developing Phase & Detecting Issues – 01/05

Second Presentation – 22/05

Final Dashboard – 07/07

Thank You

Q&A session