Ethan Gabriel Leskovec
Unimib - 886040

MotoGP

## Circuit clustering analysis

Introduction

- Dataset
- Geographical clustering
- Circuit similarity clustering
- Historical clustering
- Possible improvements
- Resources

Dataset

- The **dataset** builds upon "MotoGP circuits" by Mike Enting on Kaggle.
- It has been **modified** to add a "Short Name" column to improve the readability of the plots when hovering the mouse.
- The coordinates have been **split** into two columns to <u>improve usage</u> of the dataset.
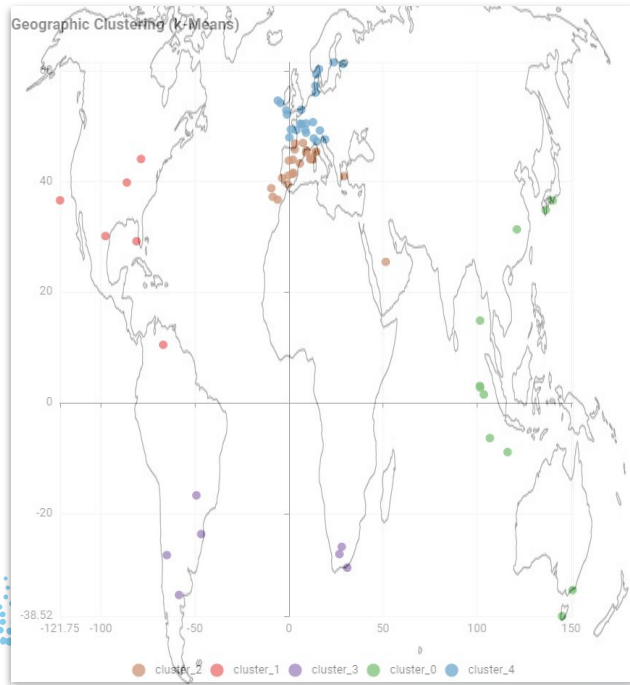
Geographic Clustering

- Clustering based on coordinates

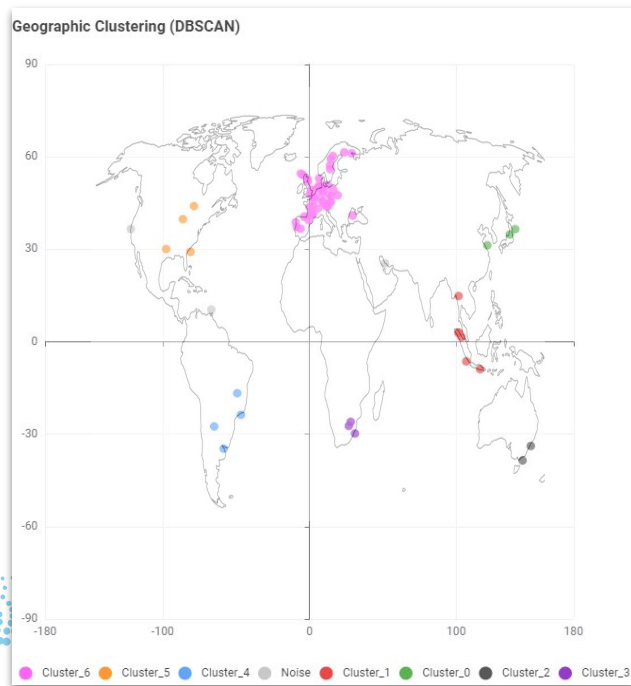# Geographic Clustering
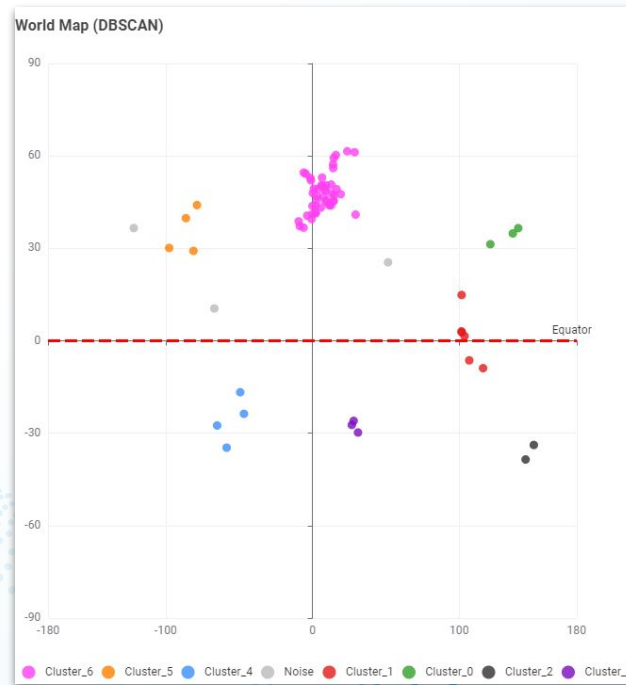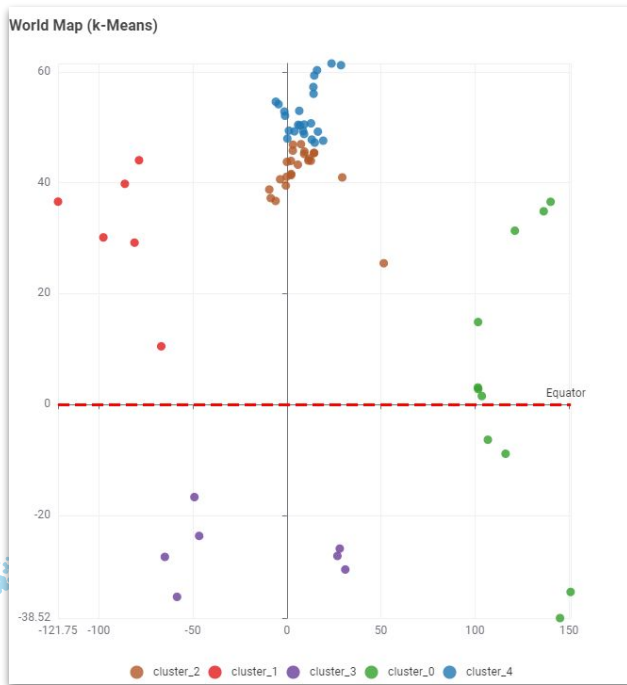
- ## k-Means

Geographic Clustering (k-Means)

k-Means uses a normalization function that has **altered** the scale of the earth, but still found the different clusters: *North and South of Europe, North America, South America (with South Africa) and Asia (with Oceania).*

Geographic Clustering

- DBSCAN

The density based algorithm has identified the different continents. There are **3 outliers** in the model: *Lusail(Qatar)*, *Laguna Seca (USA)*, *San Carlos (Venezuela)*. They are too isolated and cannot be associated with any other cluster.

6

Circuit similarity clustering

This clustering takes into consideration a couple of characteristics of the tracks (only well known tracks have been selected):
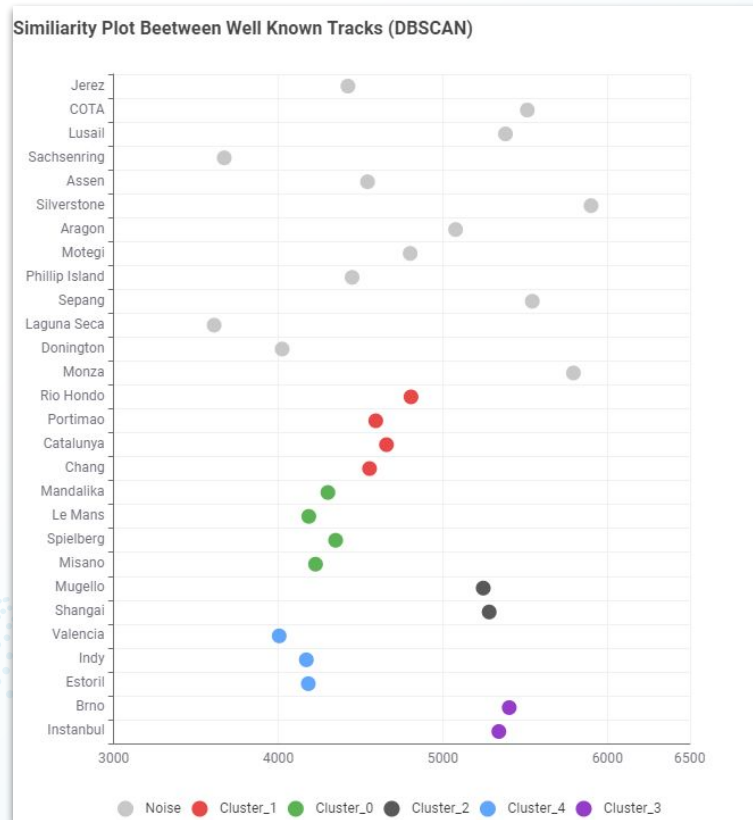
- N° of Left & Right Turns

- Avg Width

- Length

- Longest Straight

Circuit similarity clustering

- DBSCAN

As expected, a lot of **outliers** are present.
The **epsilon** and the **min. Points** parameters have been adjusted to have some reasonable clusters.

Similiarity Plot Beetween Well Known Tracks (DBSCAN)

8

Circuit similarity clustering

Example: Portimao–Catalunya

We can use this example to show how the model works. Portimao and Catalunya are often compared and are considered to be very similar tracks. The algorithm has correctly put them together.
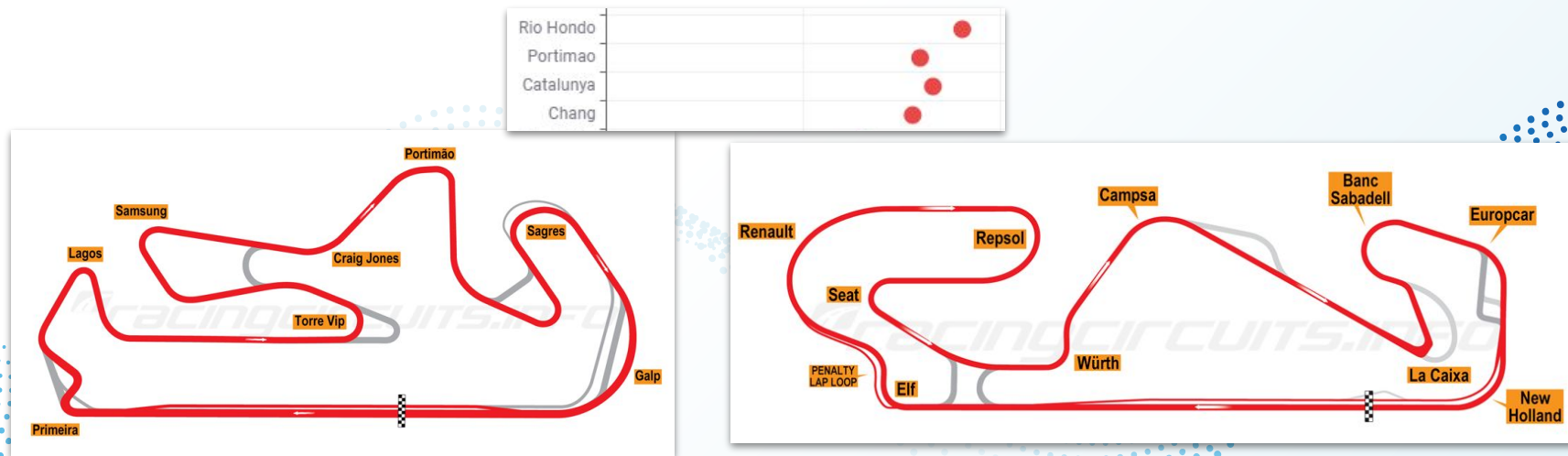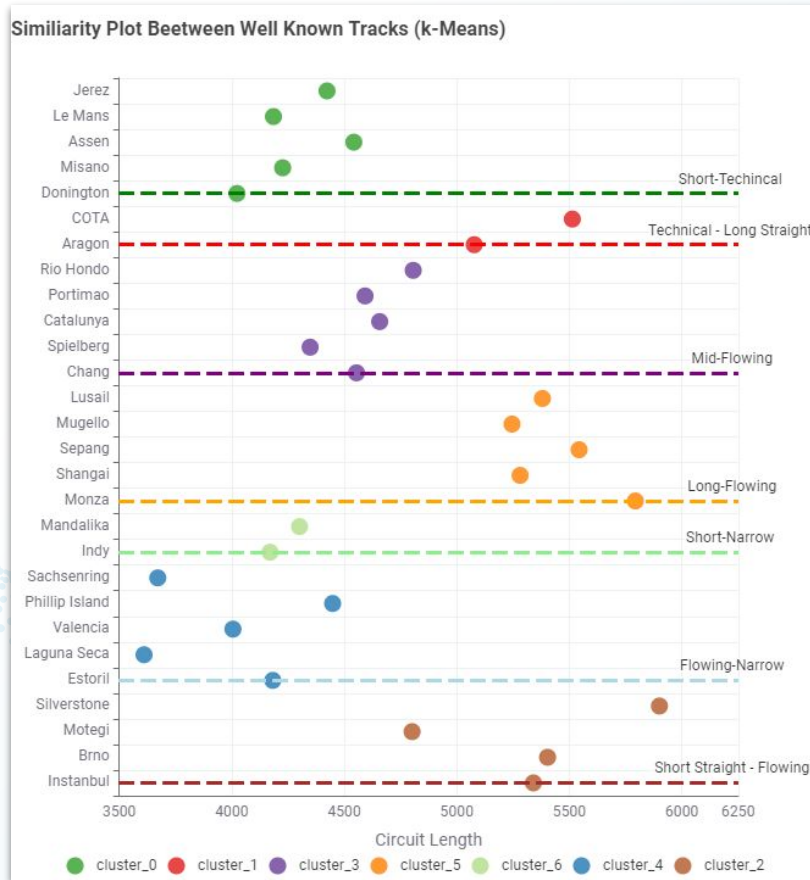
Circuit similarity clustering

- ## k-Means

The tweaking of the **"k"** parameter (for the number of clusters) has been applied here to better understand the results.

Since k-Means does **not detect outliers**, and clusters the whole dataset, captions have been added manually to have a better comprehension of the clusters.
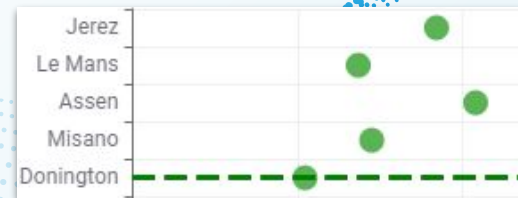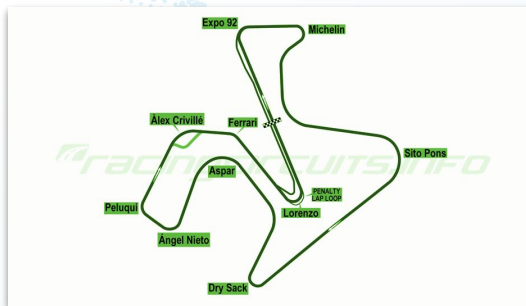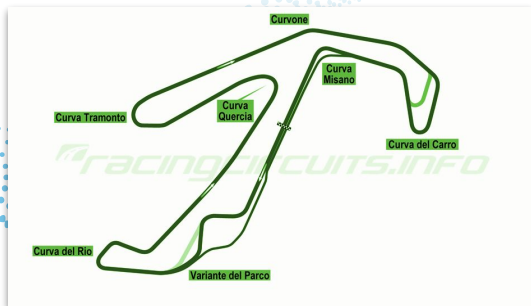

Similiarity Plot Beetween Well Known Tracks (k-Means)

10

Ethan Gabriel Leskovec
Unimib - 886040

Circuit similarity clustering

A closer look k-Means

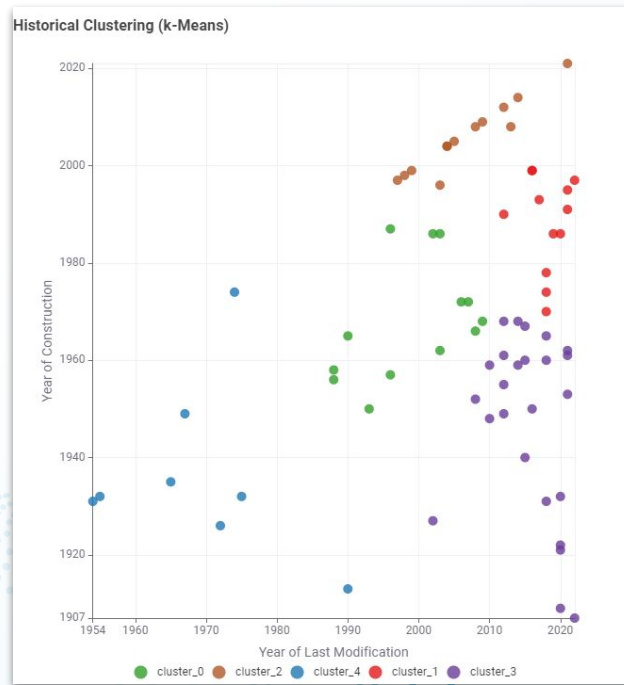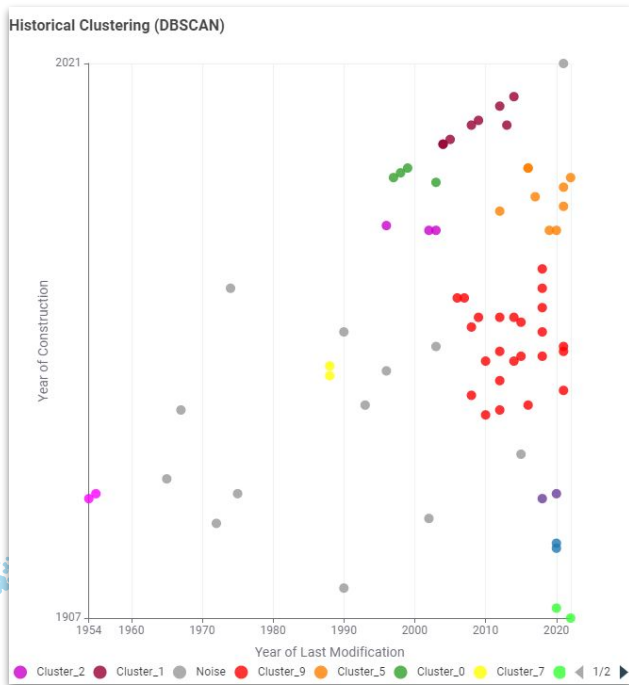**Portimao** and **Catalunya** are still best friends.

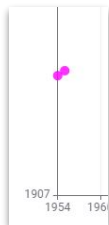**Misano** and **Jerez** is another great example.

Historical Clustering

- Clustering based on construction and modification year
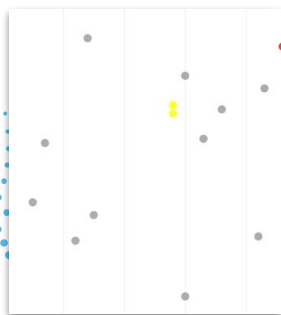
Historical Clustering

- DBSCAN highlights

Berne(1931,1954) and **Belfast(1932,1955)** are both old and now closed and unused tracks.
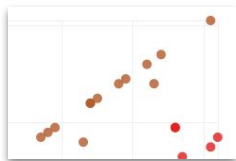

These are tracks that are more recent, like **Shanghai (2004)**, **Istanbul (2005)**, **COTA (2012)**, **Chang (2014)**, that have been built recently and have not been changed since. The **outlier** (in gray) is **Mandalika (2021).**
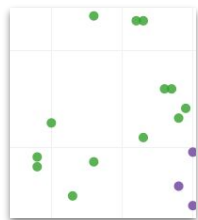

There is a lot of noise in this dataset (and not dense enough), and a **lot of tracks** have not been included in any cluster.

13

Historical Clustering

- k-Means highlights

The brown cluster has the recent, less changed tracks, **Mandalika (2021)** has been included in the cluster.

This green cluster contains older tracks, still used today, like **Suzuka(1962,2003)**, **Misano(1972,2007) and Le Mans (1966-2008)** that have not been changed in recent years.

These are **Indianapolis(1909,2020)** and **Island of Man (1907,2022)** the two oldest in the dataset. In **k-Means** they have been added to the larger purple cluster. However this is not very accurate, and only **DBSCAN** highlights their uniqueness.

14

Possible improvements

- Adding more data to the circuits:
  - Maximum Elevation change
  - Anti/Clockwise
  - Asphalt type/Kerb height/Altitude
  - Lap Record/Record Holder/Top Speed/Min Speed
- Changing distance functions
- Changing clustering algorithm parameters

Resources

- Original, pre-modification, Dataset: "MotoGP circuits" by Mike Enning on Kaggle
- Track layout pictures by RacingCircuits.info
- KNIME
- To access all the resources please visit the GitHub repository

The End