

# MotoGP

## Circuit clustering analysis



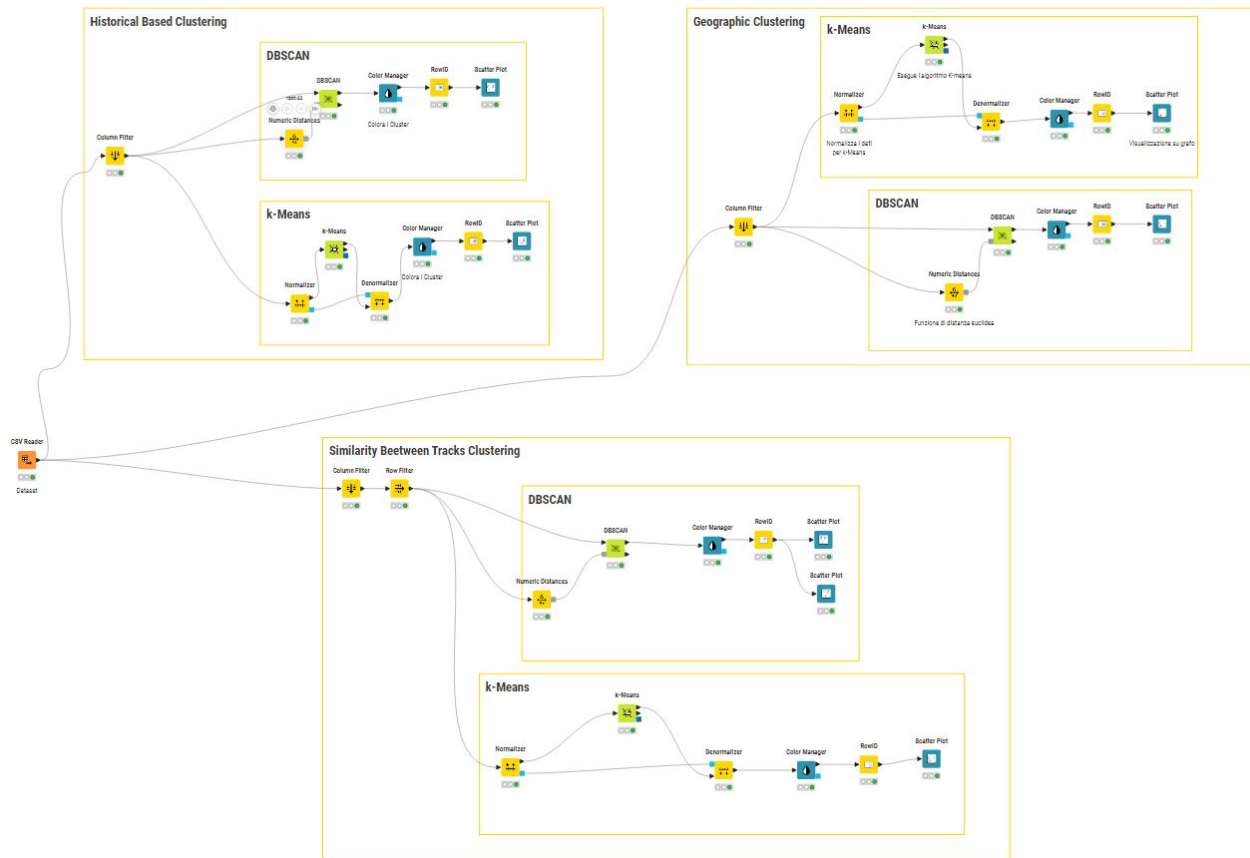
# Introduzione

- Dataset
- Clustering geografico
- Clustering basato sulla similarità
- Clustering storico
- Possibili migliorie
- Risorse

# Dataset

- Il **dataset** estende “MotoGP circuits” di Mike Enting su Kaggle.
- E' stato **modificato** aggiungendo la colonna “Short Name” per migliorare la leggibilità dei grafici quando si passa sopra con il mouse.
- Le coordinate sono state **divise** in due colonne per migliorare l'utilizzo del dataset.

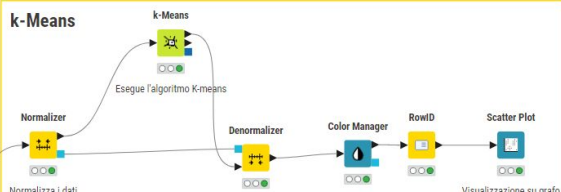
# KNIME Workflow



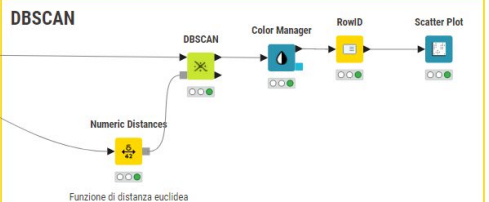
# KNIME Workflow

## Geographic Clustering

### k-Means

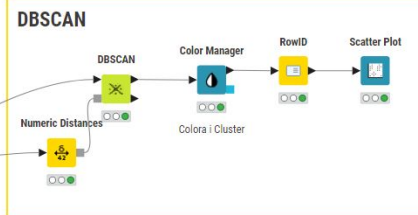


### DBSCAN

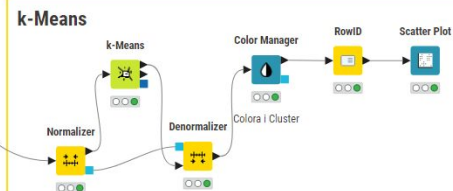


## Historical Based Clustering

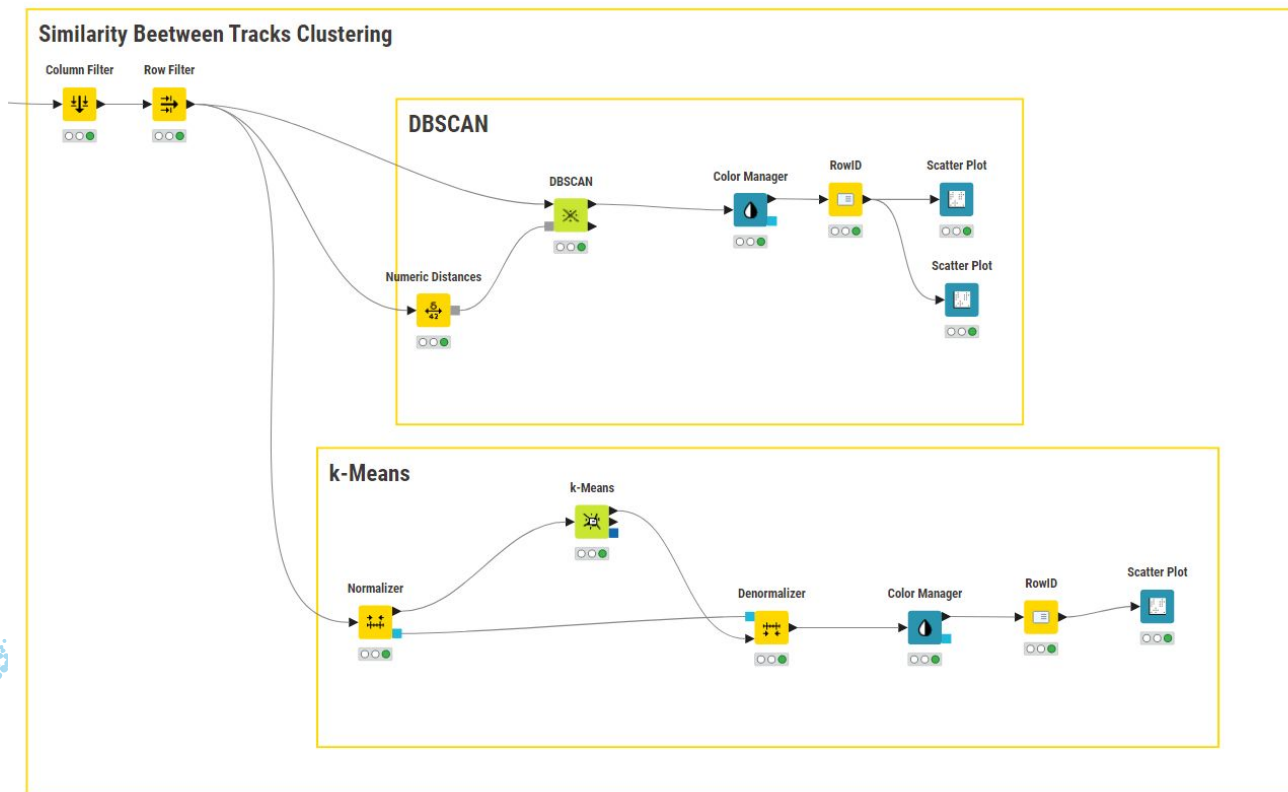
### DBSCAN



### k-Means

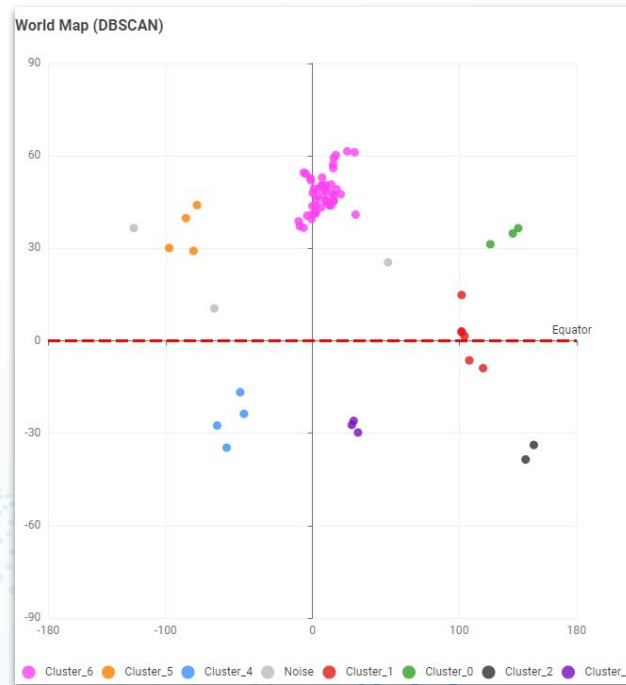
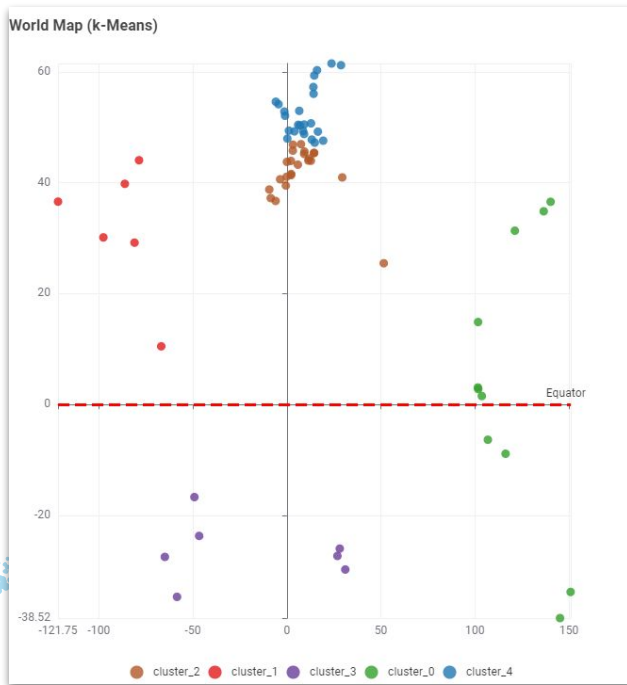


# KNIME Workflow



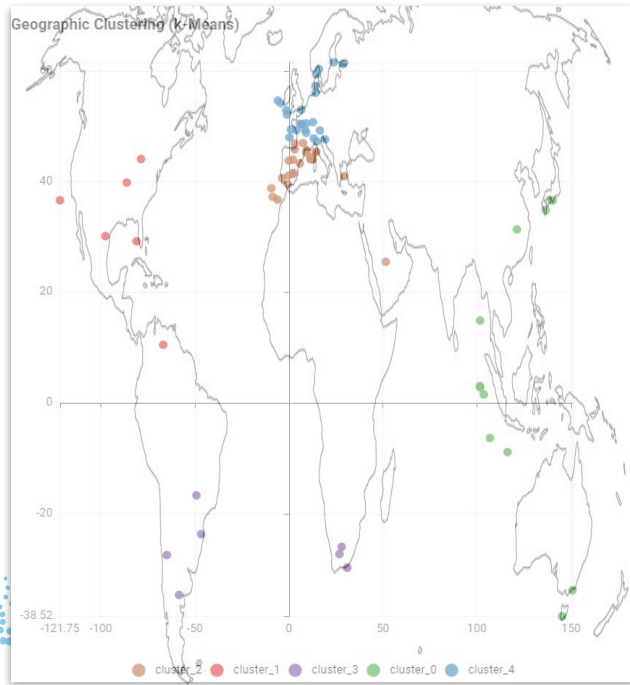
# Clustering Geografico

- Clustering basato sulle coordinate



# Clustering Geografico

- k-Means

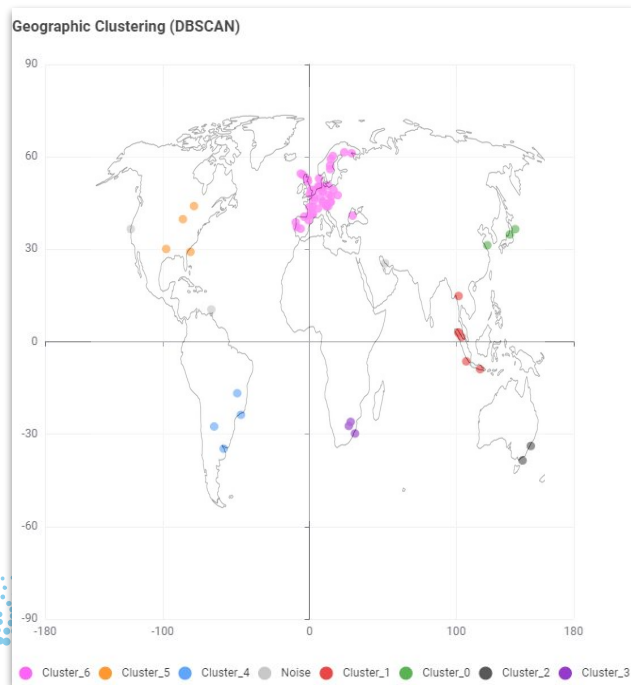


k-Means utilizza una funzione di normalizzazione, che ha **alterato** le proporzioni terrestri, ma ha comunque trovato i diversi clusters: *Europa Nord e Sud, Nord America, Sud America (con il Sud Africa) e l'Asia (con l'Oceania)*.



# Clustering Geografico

- DBSCAN



L'algoritmo basato sulla densità ha trovato i diversi continenti. Ci sono **3 outliers** nel modello: *Lusail(Qatar)*, *Laguna Seca (USA)*, *San Carlos (Venezuela)*. Sono troppo isolati e non possono essere associati a nessun'altro cluster.

# Clustering basato sulla similarità

Questo clustering considera qualche caratteristica dei circuiti (sono state selezionate solo piste ben note):

- N° di curve a Destra e Sinistra
- Larghezza media
- Lunghezza
- Rettilineo più lungo

# Clustering basato sulla similarità

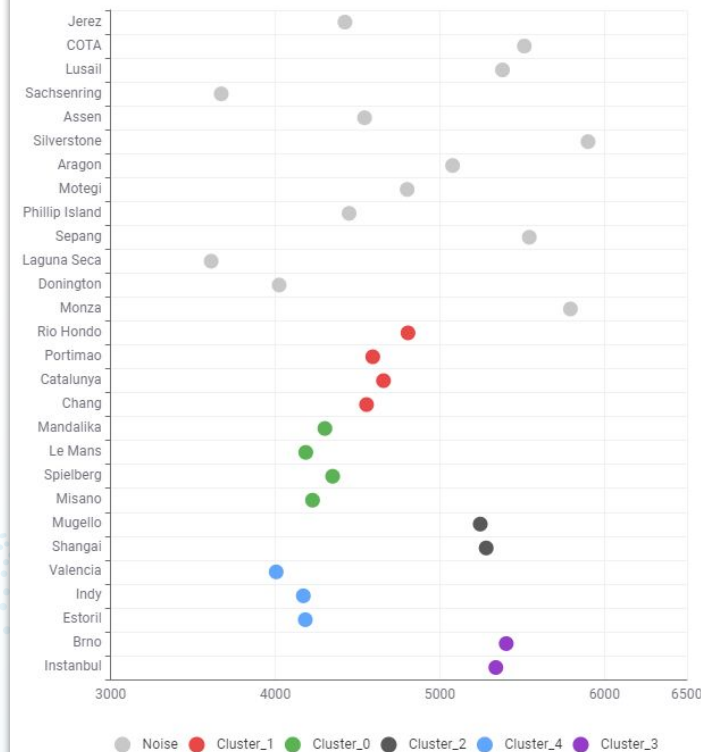
- DBSCAN

Come potevamo aspettarci,  
sono presenti **outliers**.

I parametri **epsilon** e **min.**

**Points** sono stati modificati per  
avere cluster ragionevoli.

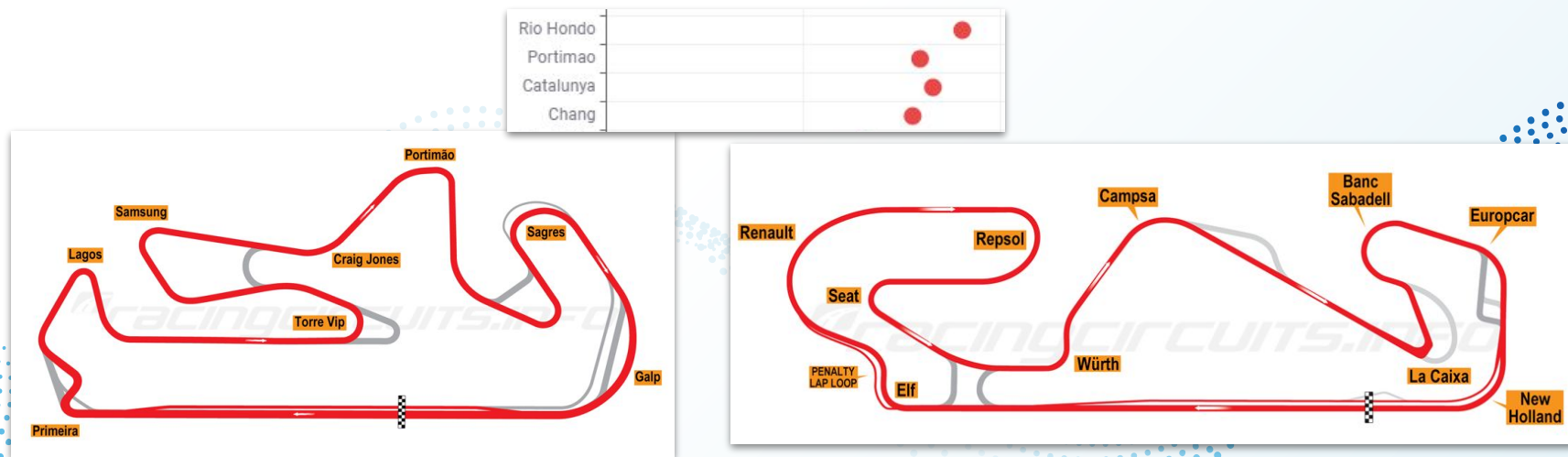
Similarity Plot Between Well Known Tracks (DBSCAN)



# Clustering basato sulla similarità

## Esempio: Portimao-Catalunya

Possiamo usare questo esempio per capire come funziona il modello. Portimao e Catalunya sono spesso comparate e sono considerate molto simili. L'algoritmo le ha correttamente inserite insieme.

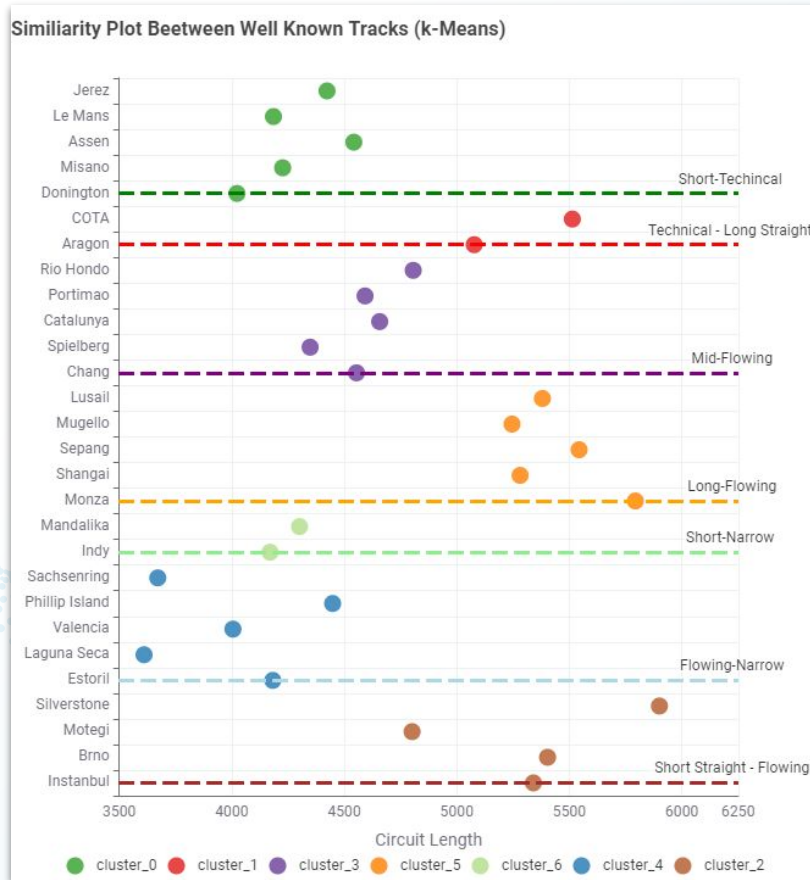


# Clustering basato sulla similarità

- k-Means

Il parametro “**k**” (numero di cluster desiderati) è stato modificato in modo da poter capire meglio i risultati.

Siccome k-Means **non individua outliers**, ma crea cluster su tutto il dataset, sono state aggiunte delle didascalie per capire meglio i cluster.



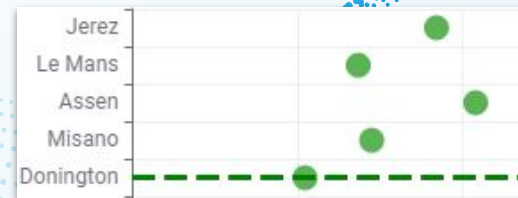
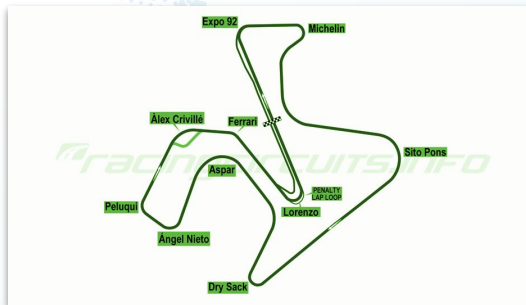
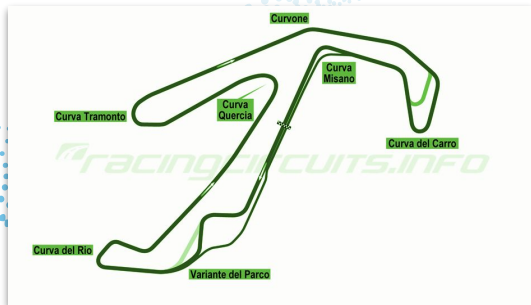
# Clustering basato sulla similarità

## Dettagli su k-Means

**Portimao** and **Catalunya** rimangono migliori amici.

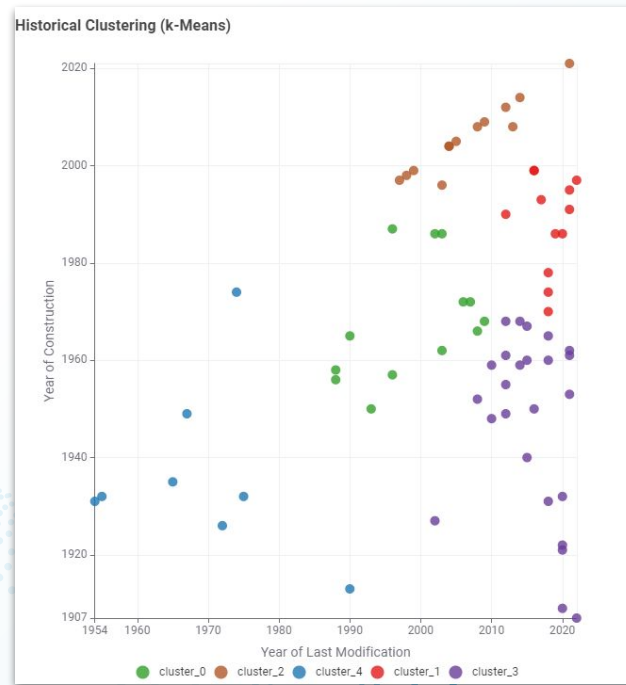
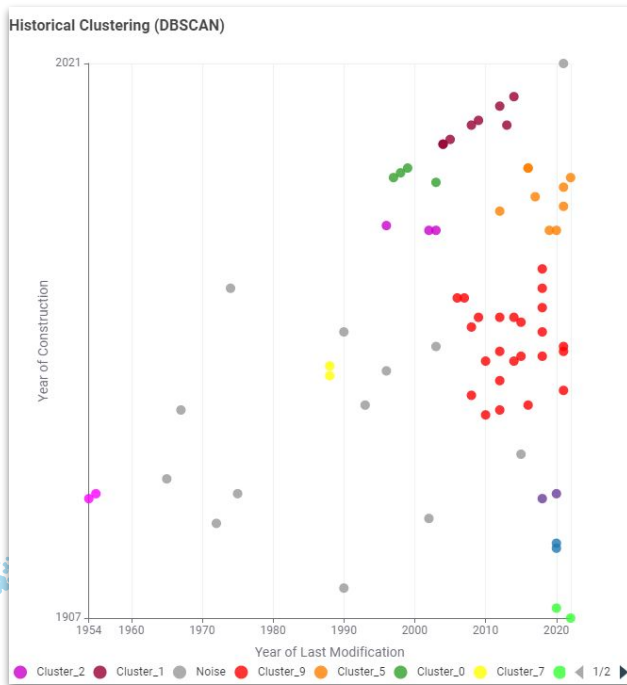


**Misano** e **Jerez** è un altro ottimo esempio.



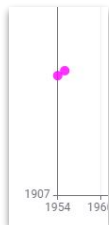
# Clustering Storico

- Clustering basato sull'anno di costruzione e di modifica



# Clustering Storico

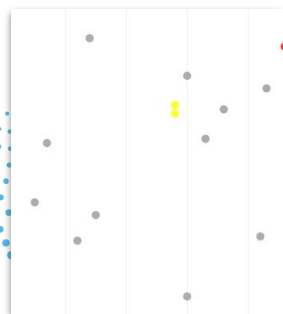
- DBSCAN highlights



**Berne(1931,1954)** e **Belfast(1932,1955)** sono circuiti entrambi vecchi e in disuso. I soli in questa condizione.



Queste piste sono più recenti, come **Shanghai (2004)**, **Istanbul (2005)**, **COTA (2012)**, **Chang (2014)**, che sono state costruite recentemente e quindi non sono state modificate da allora. L'**outlier** (in grigio) è **Mandalika (2021)**.

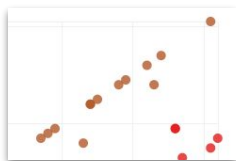


Questo dataset è parecchio rumoroso (e non abbastanza denso), e **molti circuiti** non sono stati inseriti in nessun cluster.

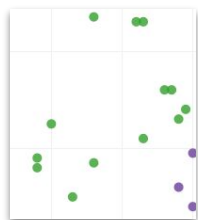


# Clustering Storico

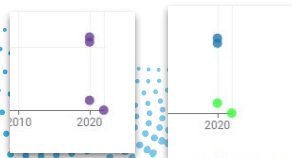
- k-Means highlights



Il cluster marrone contiene le piste recenti, meno modificate, **Mandalika (2021)** è stata inclusa nel cluster.



Il cluster verde contiene piste più vecchie, ma ancora in uso, come **Suzuka(1962,2003)**, **Misano(1972,2007)** e **Le Mans (1966-2008)** che non hanno subito modifiche di recente.



Queste sono **Indianapolis(1909,2020)** e **Island of Man (1907,2022)** le due più vecchie nel dataset. In **k-Means** sono state aggiunte nel cluster viola. Ma questo non è molto preciso e solo **DBSCAN** riesce a mettere in risalto la loro unicità.

# Possibili miglioramenti

- Aggiungere più dati ai circuiti:
  - Cambio di elevazione massimo
  - Senso Anti/Orario
  - Tipo di asfalto/Altezza cordoli/Altitudine
  - Record sul Giro/Recordman/Vel. Max/Vel. Min
- Modifica delle funzioni di distanza
- Modifica dei parametri degli algoritmi di clustering
- Aggiunta di algoritmi basati su distribuzione e gerarchia
- Espansione del dataset

# Risorse

- Dataset originale, pre-modifica: ["MotoGP circuits" di Mike Enning su Kaggle](#)
- Immagini delle piste: [RacingCircuits.info](#)
- KNIME
- Per accedere a tutte le risorse visita la [repository di GitHub](#)

# Fine

A decorative graphic consisting of multiple parallel, wavy lines of blue dots. The dots are arranged in a way that creates a sense of depth and movement, flowing from the bottom left towards the top right, framing the central text.