Customer Clustering Analytics

Introduction

What makes a data scientist outstanding is the hability to convert data into valuable and actionable insigths and to provide bussines added value. How awesome your Python coding or modeling skills can be, these may not be enough to be a good data scientist but the valuable insigths you can get from data and present them in a clear way.

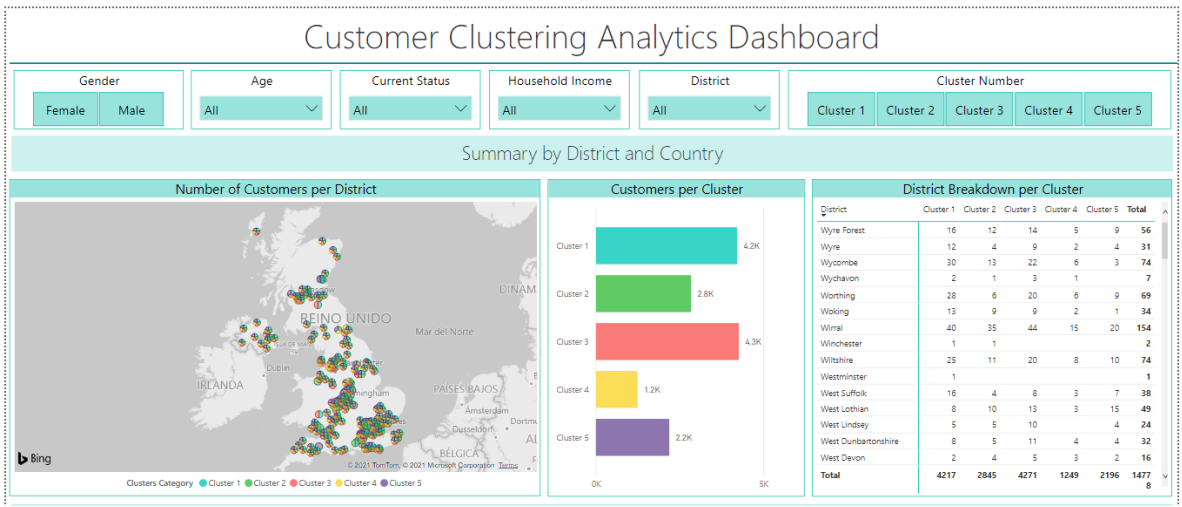
Machine learning is the process of programming computers in order they can learn from data. We use statistical models and alogithms to perform tasks like classification and prediction without explicit instructions. This technology can be apply to help companies or organizations that provide a service or product to let them know their customers, how they behave and how the company can increase their impact within their customers and increase revenue.

Objetive

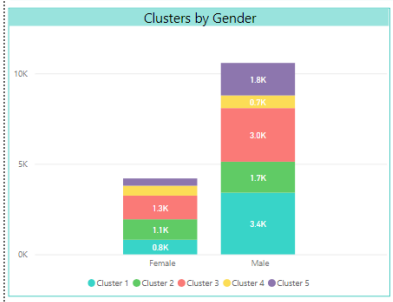
Our aim is to identify groups of data with similar characteristics or behaviours within a customer dataset. Some of the data that includes this dataset are variables like Age, Gender, District, Household income, Status and information that surrounds events attending. We want investigate what are the common traits among each cluster and deploy our results into a Power BI dashboard for the bussines to use and add them value.

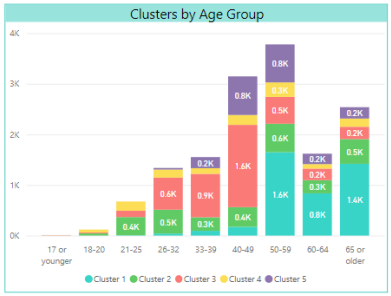
We first need Python to preprocess the data and apply the K means algorithm to the data and then use this data output we get to build a dashboard in Power BI.

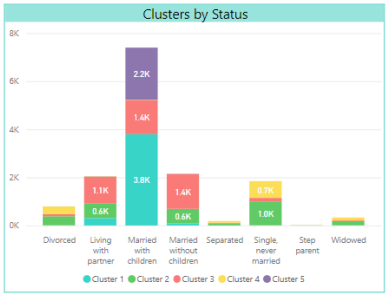
Dashboards

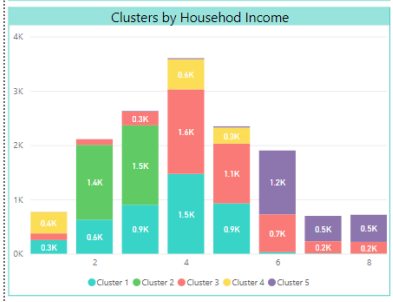


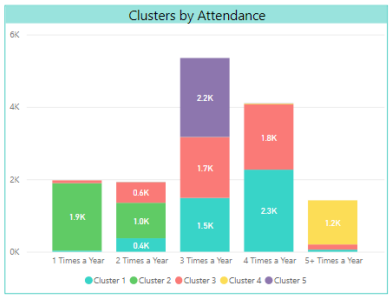
Clusters Features Analysis

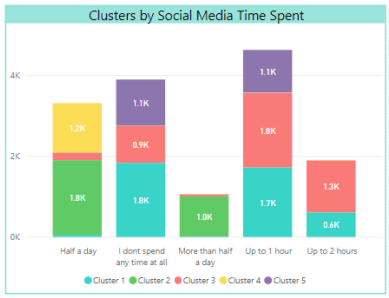


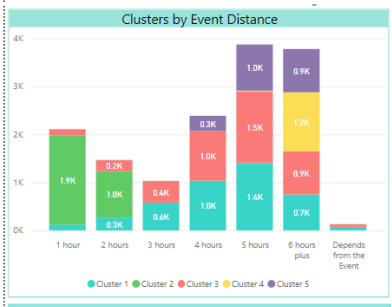


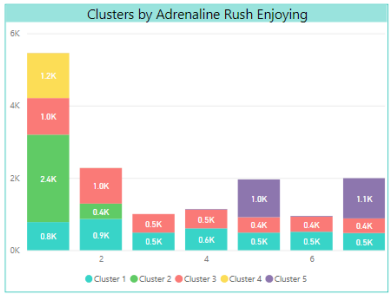


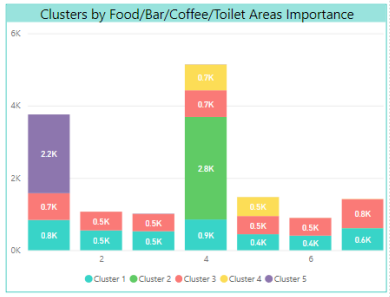


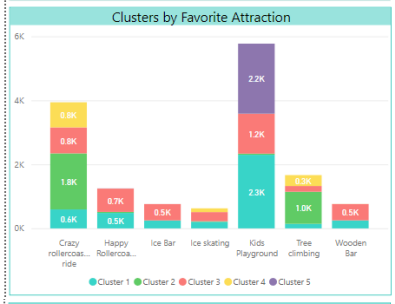


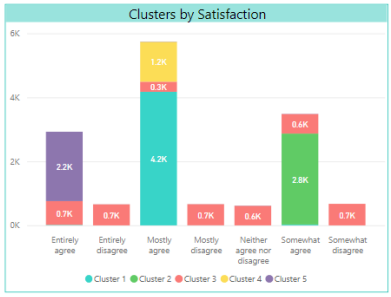


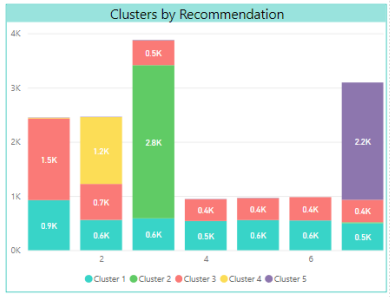


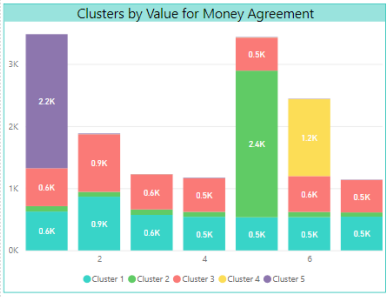


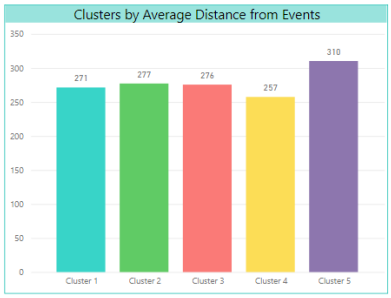


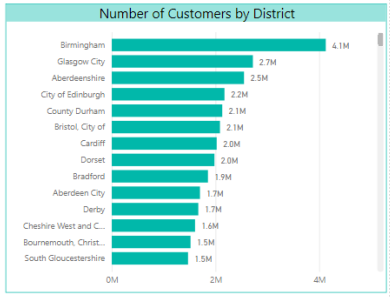












Cluster 1 Features

- Mostly people with age being 50+
- Mostly married with children
- Household income ranges from 25k to 100k
- Attend events 3 to 4 times per year
- Don't spend too much time on social media (< 1 hour)
- Kids playgrounds is their favorite attraction

Cluster 2 Features

- People who don't have kids - mostly single
- Earn between 25k to 50k
- Attend events mostly once or twice a year
- Spend lot of time in social media (half a day+)
- Love adrenaline rush activities
- Not bothered with food/coffee/bars/toilet areas

Cluster 3 Features

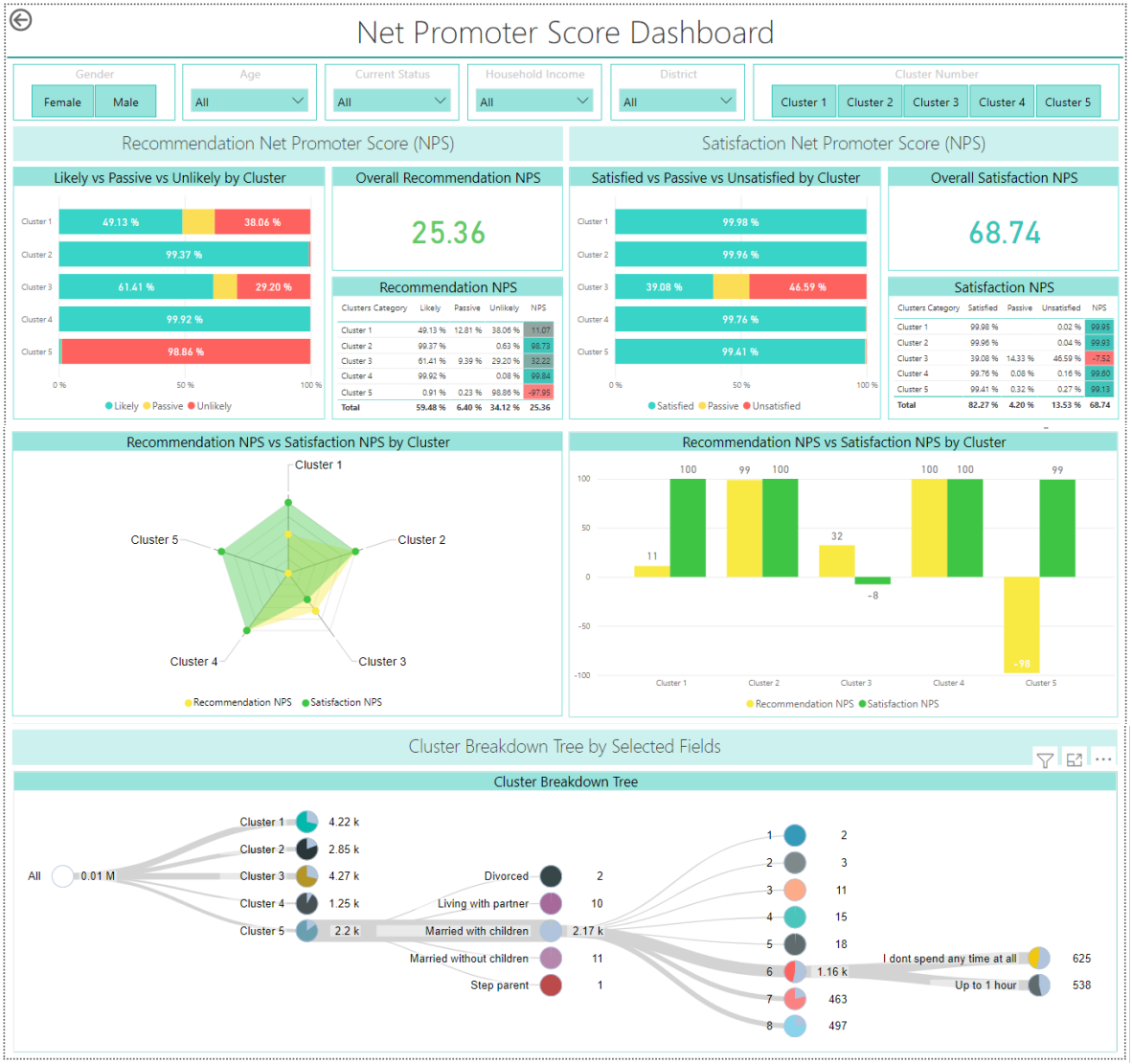
- Mostly people with age range between 26 to 50
- Married people who have kids or living with their partners (2+)
- Earn between 50k to 150k
- Attend events 3 to 4 times a year
- Spend between 1 to 2 hours in social media
- Mostly willing to travel 4 to 6

Cluster 4 Features

- People who are single, separated, divorced or widowed
- Household income ranges between 50k to 100k or less than 20k
- Attend a lot of events per year (5+)
- Spend half a day in social media
- Willing to travel up to 6

Cluster 5 Features

- Mostly people with age between 40 to 60
- Married with children
- High earners with 100k+
- Attend events 3 times per year
- Do not spend much time on social media (1 hour)
- Willing to travel 4-6 hours for the event
- Not adrenaline people



Conclusion

Customer clustering is the segmentation of a market into discrete customer groups that share similar behavior and it can have powerful means to identify unsatisfied customer needs. Using machine learning, we can get insigths that otherwise we wouldn't be able to get. This way, organizations can outperform the competition by developing uniquely appealing strategies or products and consequently increase the revenue.

K means algorithm is an awesome option to get the job done since it allows us to explore the data to a very specific level. With this Project we were able to have a sight of customer distribution by every variable and go further by analyzing the net promoter score that measures customer satisfaction and an indicator of the growth of the company.