Appendix: Emma Luk (Mobile: 07974 522 805): A Selection of my data science work

1A: Deep Learning with TensorFlow Long Short-Term Memory (LSTM) Neural Network for Stock Market Predictions with Python

Figure 1: Predicted stock prices and Actual Stock Prices

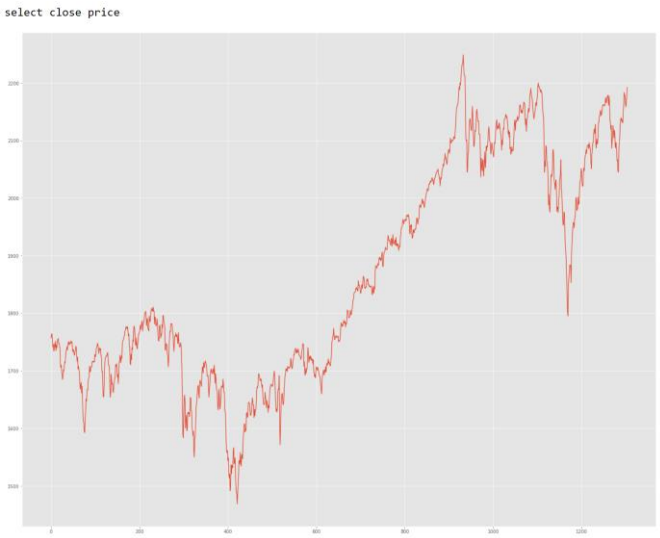
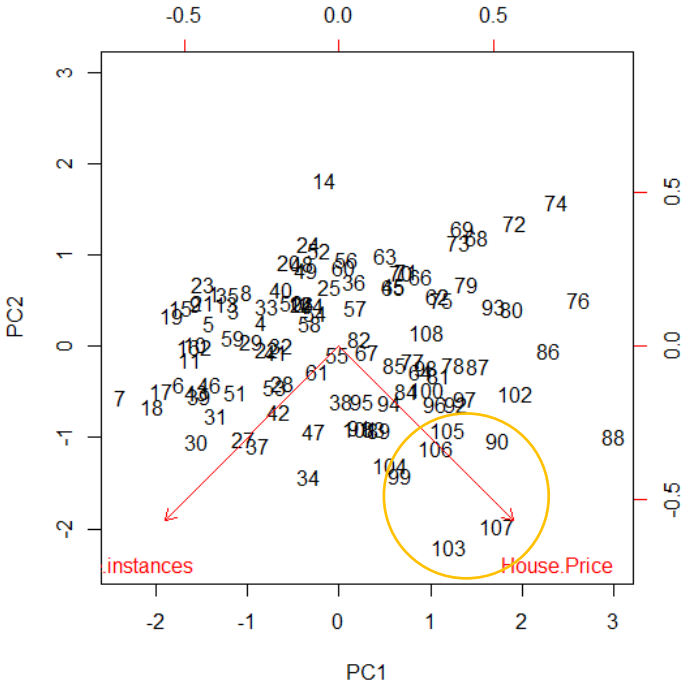


Figure 2: Plot of 'Close' for Global Equity Income sector price history

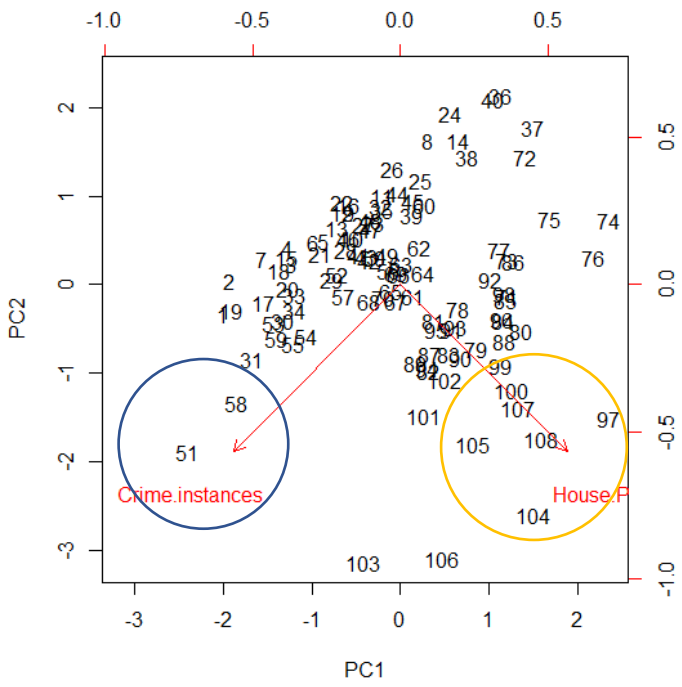
1B: Analysing House Prices and Crime in R

Main analysis Unsupervised learning: Principal Component Analysis (PCA):

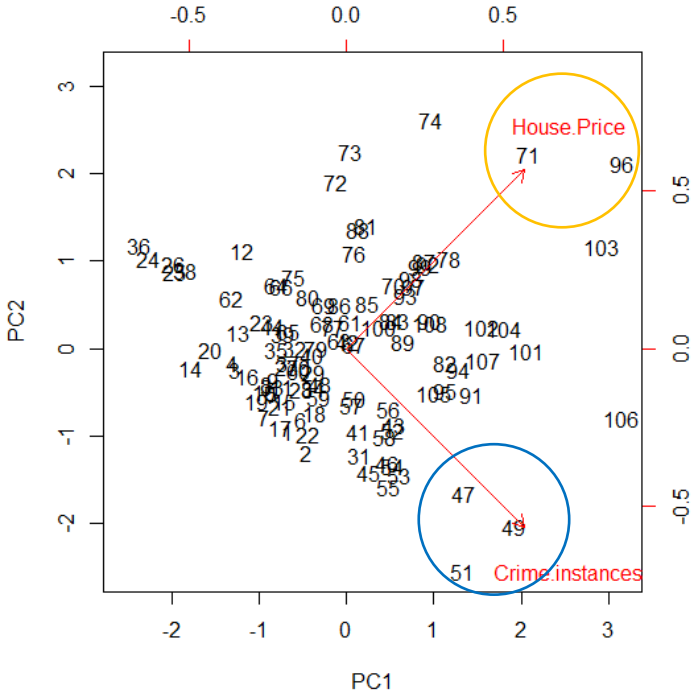
- The aim was to investigate how current house prices were affected by recent crime levels in London Boroughs
- Performed data cleaning, transformation, manipulation and conducted Principal Component Analysis (PCA): this was the process by which compute principal components and used them for **better understanding of the data**. PCA is considered an **unsupervised machine learning** method because it involves only a set of feature variables and no associated response variable. PCA also serves as a useful tool for **exploratory analysis** and **data visualisation**



Merton



Wandsworth



Hounslow

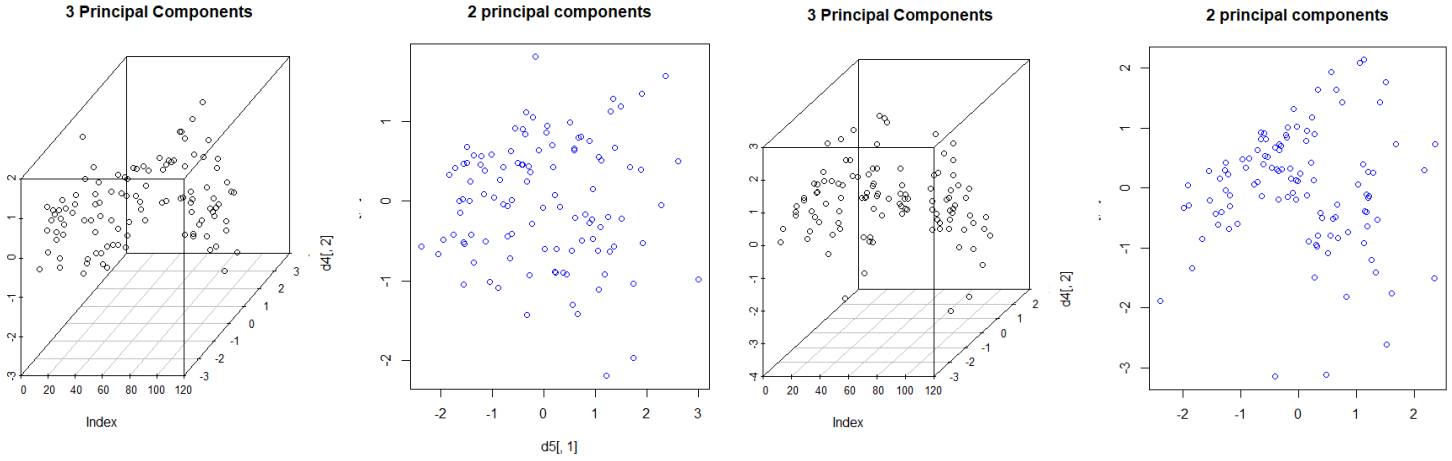
In Wandsworth:
House 51: £ 538,999.3
House 104: £ 973,938.4

In Hounslow:
House 71: £689,162.6
House 49: £396,623.0

For PCA:
House 104 & 71 data points are near House Price point. They are more expensive.

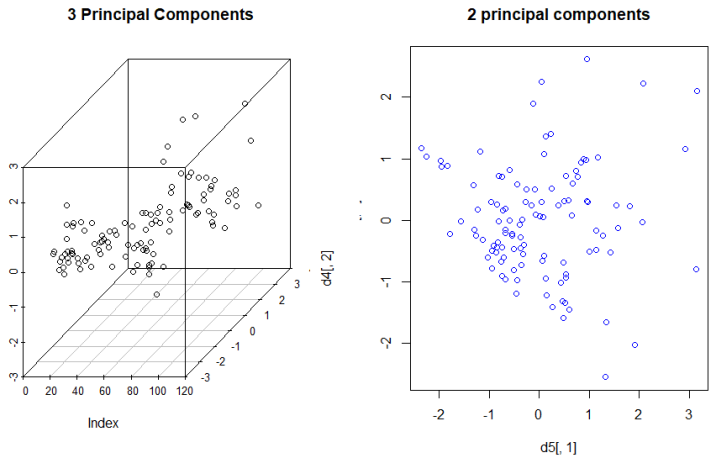
House 51 and 49 data points are near Crime point. They are less expensive. This is an indication that crime has some effect on house prices.

Figure 3: Plot of the first two principal components (PC1 & PC2) for Hounslow, Merton & Wandsworth



Merton

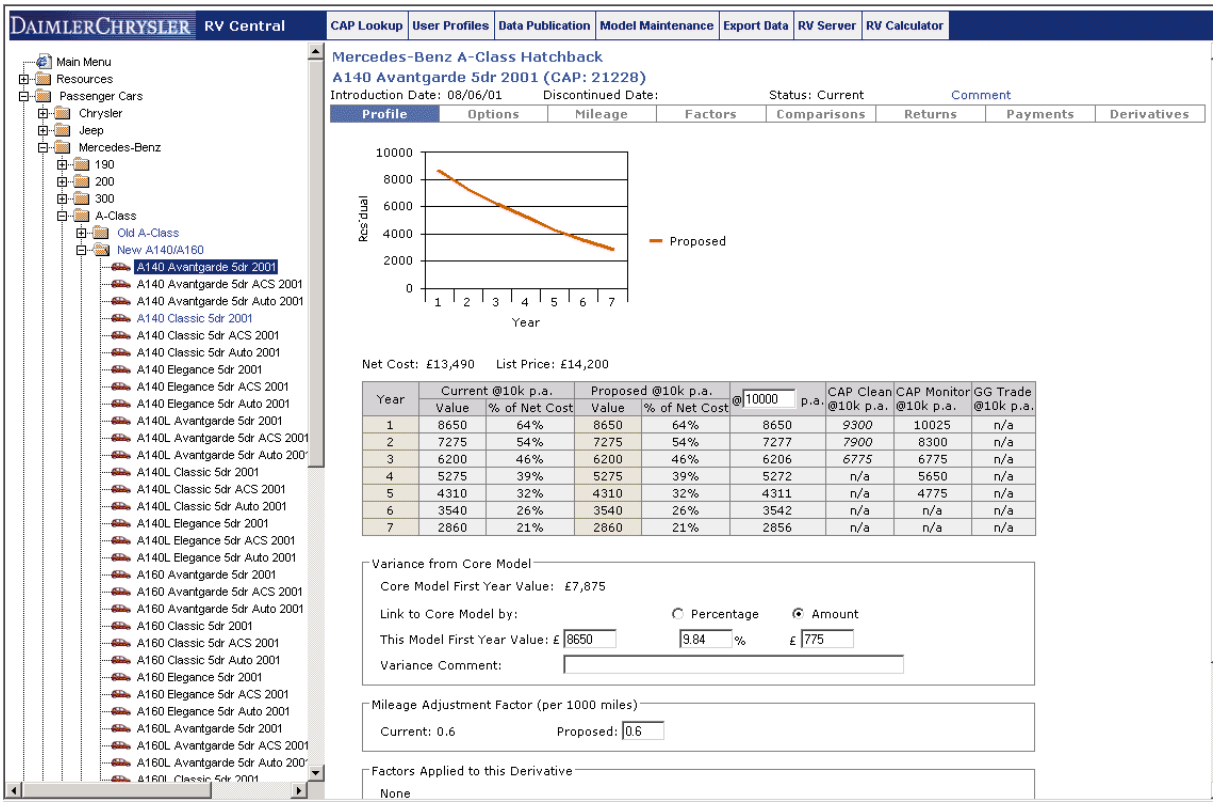
Wandsworth



Hounslow

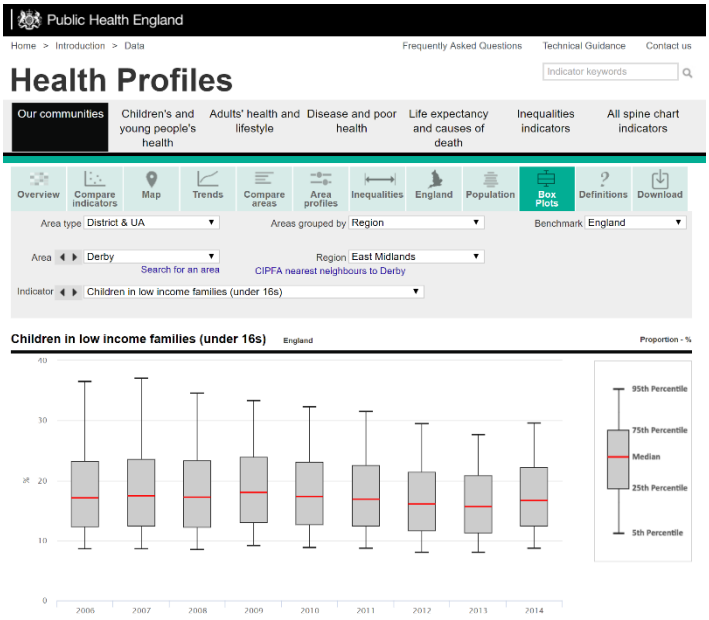
Left: Figure 4: 3D scatterplot using three principal components
Right: Figure 5: 2D scatterplot using two principal components

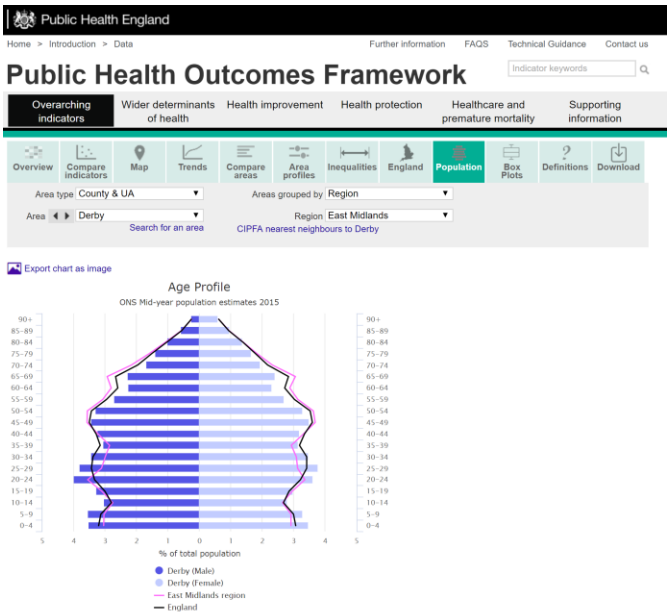
1C. An example from when I worked at Mercedes-Benz UK: The first ‘RV Central’ in the UK, which provides the residual value of your used car. This system calculates the residual value for used cars and has since been implemented by Mercedes-Benz throughout all their UK dealerships. I also applied data science skills and new technology to improve business processes and business efficiency and reduced costs; this helped the company implement automated processes that would previously have been paper-based, resulting in increased efficiency and reduced environmental impact.



1D. Examples of A/B Testing and Multivariate Testing (MVT) from when I worked at AXA. The two examples below depict different versions of the same webpage, which were used to provide insight to drive future strategies and identify business opportunities and problems.

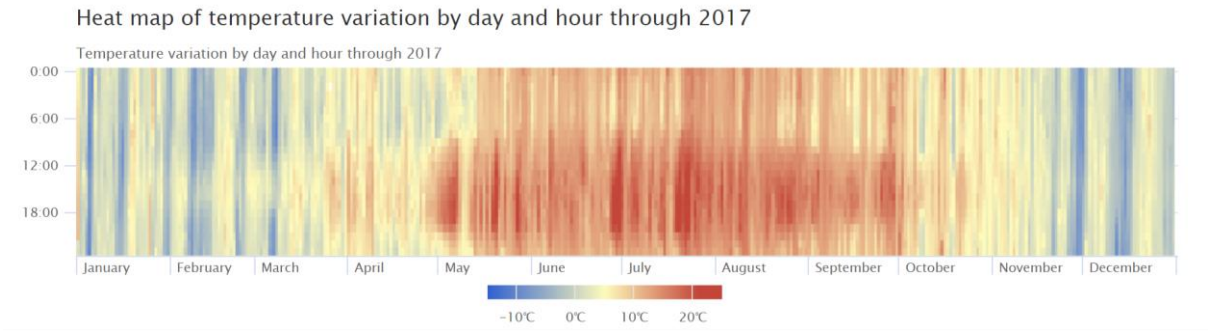
1E. Examples of data visualisation from when I worked on the 'Fingertips' website for Public Health England (Government Agency): On the left are boxplots depicting the percentage of children in low income families in the East Midlands between the years 2006 and 2014; on the right is a bar graph with negative stack depicting the proportion of males and females of different age groups in the East Midlands region.





1F Industry Project: How Robots are making Farming Profitable
Weather Data Analytics Using Hadoop

- Leading the big data flow of the application starting from data ingestion from upstream to HDFS, processing and analysing the data in HDFS and data visualisation in R & JavaScript



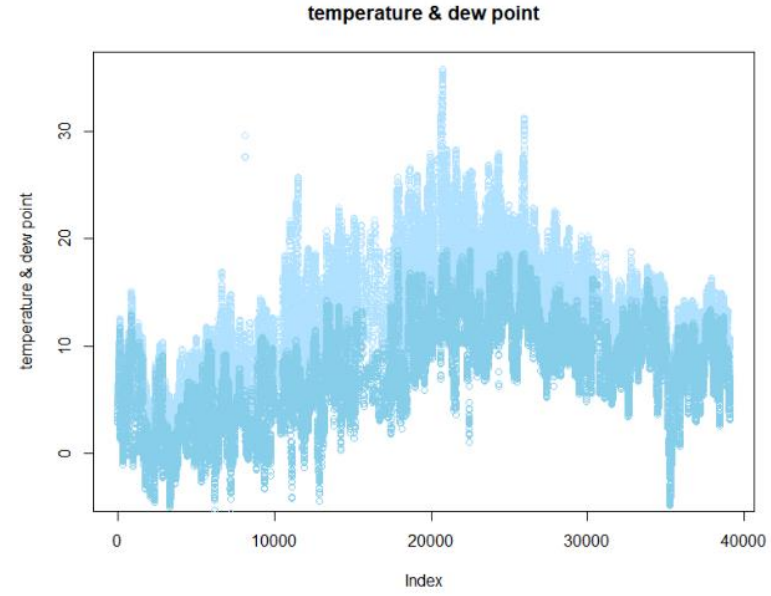
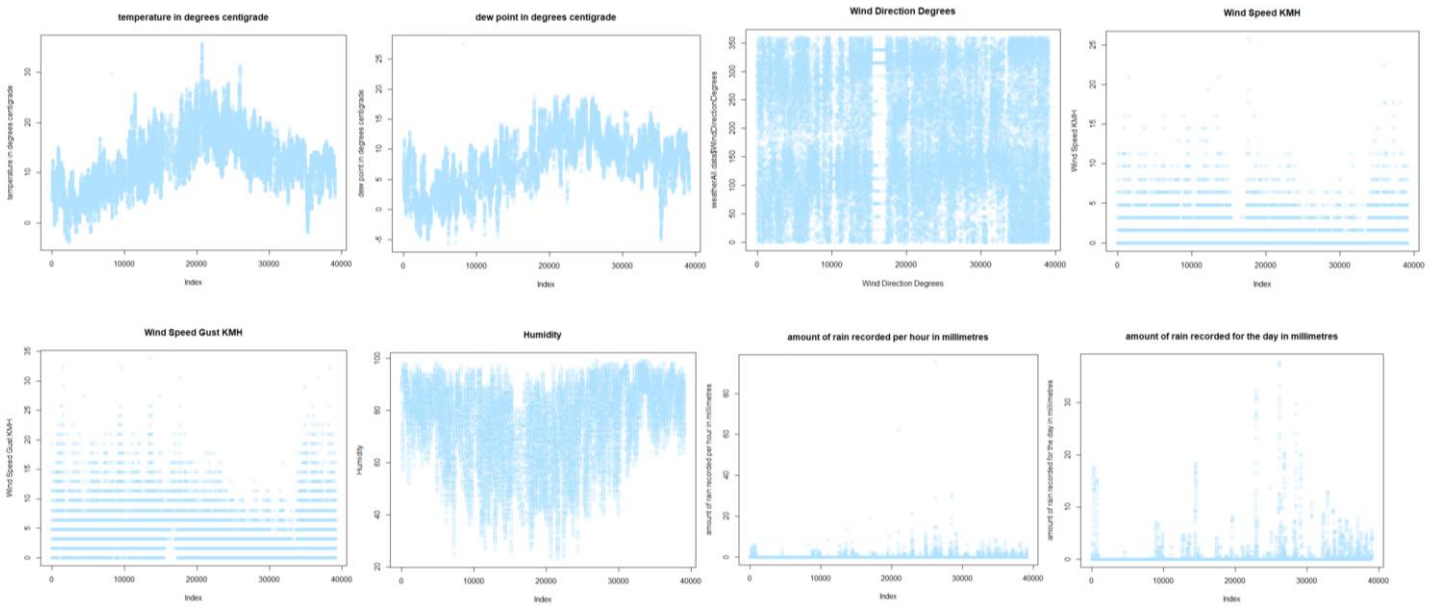


Figure 6: Scatter plots with two level
Added two level first level is dew point and second level is temperature in R



1G: Data Visualisation in Microsoft Power BI & Tableau

- The aim was to create Executive Dashboard, tracked and reported on business metrics & the KPIs
- This dashboard included key performance, top ten products, top performing cities and top performing city, customer reviews and sales by Month.

Microsoft Power BI:

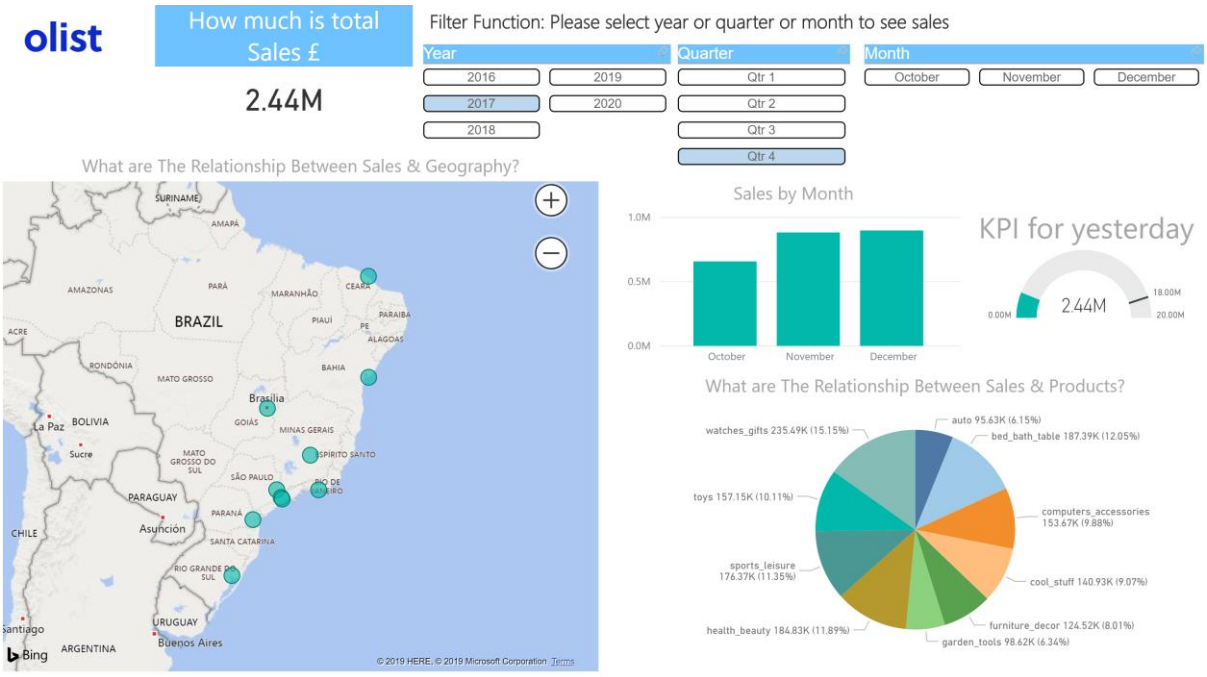


Tableau: