Emma Luk

Address: Guildford, UK, Mobile: 07974 522 805 Email: emma_luk2@hotmail.com LinkedIn Data Science Projects

Summary

- Studied MSc in Data Science and Analytics to further develop skill sets in **machine learning, big data & visual analytics** technologies. **Industry knowledge in the healthcare, financial, automotive and insurance sectors with Data Security**
- Worked as a Web Designer / Developer for more than fifteen years. Gaining valuable IT and client liaison skills
- Offer extensive international experience either leading a team or individually undertaking event-based **analytics, predictive modelling, Google Analytics, A/B testing, build business modelling, testing and implementation**
- Pride myself on producing results because I effectively **plan ahead, incisively identify potential problems, organise time and resources and follow through passionately to ensure that results are achieved**
- At Public Health England (Government Agency) (**Appendix 1E**), where applied **data science and data visualisation** for 'Fingertips', a rich source of indicators across a range of health for Public Health Intelligence
- Whilst at AXA (**Appendix 1D**), I applied machine learning, in building models **that translate data points into business insights and algorithms to improve the accuracy to monitor online customer behaviours, identify problems and opportunities, and gather insight for future strategies**. **Involved in reporting, risk analysis, compliance & integrations**
- At Mercedes-Benz UK (**Appendix 1C**) worked as part of a team that innovated and created the first 'RV Central' in the UK. This helped the company **implement automated processes that would previously have been paper-based, resulting in increased efficiency and reduced environmental impact**
- Tech Stack: Python, Jupyter Notebook, Statistics with R, RStudio, NoSQL, MS SQL, HTML

Education and Professional Certification

2018 – 2020 **MSc Data Science and Analytics** Brunel University, London

Key modules

- **High Performance Computational Infrastructures in Java** (Hadoop & NoSQL)
- **Data Visualisation** (**dashboard and infographic design using Tableau & Microsoft Power BI**).
- **Big Data Analytics** covers data mining alongside machine learning techniques (e.g. clustering, regression, support vector machines, boosting, decision trees and neural networks) etc
- **Quantitative Data Analysis & Statistics with R**

MSc - specialising in **machine learning, big data and visual analytics technologies**. Experienced at programming in Python with Jupyter Notebook, Hadoop, Spark, R with RStudio, NoSQL, Java and Unix systems.

Dissertation - Next Generation Fund Managers: Deep Learning with TensorFlow **Long Short-Term Memory (LSTM)** Neural Network (Recurrent Neural Network) for Stock Market Predictions with Python **(Appendix 1A)**.

2017	**Certified Scrum Master® (Credential ID: 000691196)**	Scrum Alliance
2006	**MSc in Computer Science**	University of Bristol
1997	**Post-graduate Diploma in Multimedia Design**	Staffordshire University
1994	**BA (Hons) in Graphic Design (2:2)**	University of Plymouth

Data Science Projects

Deep Learning with TensorFlow LSTM Neural Network for Stock Market Predictions with Python (Appendix 1A)

- The project used the TensorFlow and Keras libraries to implement a LSTM neural network
- Developed Sequence Prediction Models with Deep Learning for Stock Market Predictions
- Worked extensively with Neural Network and deep learning techniques for Stock Market Predictions

Completed the following tasks to build the model:

1. Select problem
2. Getting Data
3. Setup a Python Environment for Machine Learning and Deep Learning
4. Reading and analysing data using Pandas
5. Normalising the data using SkLearn
6. Converting data to time-series and supervised learning problem.
7. Creating model
8. Fine tuning the model
9. Training, predicting and visualising the result.

Industry Project: How Robots are making Farming Profitable

Weather Data Analytics Using Hadoop (Appendix 1F)

- Leading the big data flow of the application starting from data ingestion from upstream to HDFS, processing and analysing the data in Hadoop Distributed File System (HDFS) and **data visualisation** in R

Analysing House Prices and Crime in R (Appendix 1B)

- 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**

Business Insights at AXA (Appendix 1D)

- Devising and implementing **effective tracking metrics to monitor online customer behaviours, identify problems and opportunities, and gather insight for future strategies**
- Building models that **translate data points into business insights so AXA can better understand customer behaviours**
- Used Machine Learning to accelerate A/B Testing and Multivariate Testing and helped develop algorithms to improve the accuracy of Recommendation Engines. This helped provide insight to drive future strategies and identify business opportunities and problems
- Generated 30% in incremental revenue by predicting an optimal time when a customer should be engaged with an upsell opportunity

Investigate and explore the relationships between weather conditions and energy consumption (Appendix 1G)

- The goal was to draw a graph that shows how the samples are related (or not related) to each other
- What was the relationships between weather conditions and energy consumption in London?
- PCA was used and deterministic. So, the correct answer is guaranteed. It makes data plottable on a 2D graph

Data Visualisation in Microsoft Power BI & Tableau (Appendix 1H)

- 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

---

Employment History

2020 – 2020 Data Scientist with Deep Learning (Contract), Massive Analytic Ltd, London

- Working with Deep Learning, TensorFlow, Long Short Term Memory Networks (LSTM), Fuzzy Time Series
- Developing Sequence Prediction Models with Deep Learning for Self-driving cars, unanimous data science and deep learning to build a control unit (**Appendix 1I**)
- Creating mock-ups & Prototyping

2016 – 2018 Software Developer (Contract), Public Health England (PHE) Department of Health, Cambridge

- https://fingertips.phe.org.uk/profile/health-profiles (Go to the data)
- Applied data science and data visualisation for 'Fingertips', a rich source of indicators across a range of health and data sources commonly used in Public Health Intelligence (**Appendix 1E**) with the **Government Digital Service (GDS) standards and guidelines**
- Responsible for using **data visualisation techniques** to increase the usability and accessibility of data

2014 – 2015 Web Developer (Contract), SelectScience (awarded The Queen's Awards for Enterprise International Trade)

- Designing, developing and delivering the SelectScience website using N-Tier Architecture, comprising Presentation, Business and Data Access Layers, whilst also using C#.NET, ASP.NET, CSS3 and HTML 5, BootStrap, iOS and Android

2013 – 2014 ASP.NET Developer (Contract), NMR Information Partnerships Plc, Chippenham

- Translating complex business requirements and **"Data-driven business" into tangible technical solutions**

2012 – 2013 User Experience Developer (Contract), Hospitality Guest Experience Management (Microsoft Partners)

- Designing, developing and implementing new websites, **online feedback tools** and landing pages using ASP.NET web forms, MVC 4 technology, iOS and Android

2010 – 2012 Online Marketing Webmaster, AXA Wealth, Bristol

- Conducted Machine Learning to accelerate A/B Testing and Multivariate Testing (**Appendix 1D**) and helped develop algorithms to improve the accuracy of Recommendation Engines
- This helped provide insight to drive future strategies and **identify business opportunities, infer marketing insights, forecast sales and problems,** and accounting standards and guidelines.

2004 – 2010 Web Designer / Developer, Hargreaves Lansdown Stockbrokers & Asset Management Ltd (UK FTSE 100 Company)

- Techniques have included event-based analytics, predictive modelling, Google Analytics, A/B testing, **tracking the customer experience, correlation between financial keywords Internet search and stock market fluctuations, time-series analysis**, and attribution strategy development using Waterfall and Agile processes

2001 – 2004 Web Designer / Developer, V-ten Limited (a subsidiary of DaimlerChrysler UK), Bristol

- Designed and managed the Mercedes-Benz and Smart websites (**Appendix 1C**), running projects up to the value of £4 million per year, E-commerce disciplines, a Configurator System, Content Management System and database implementation, and accounting standards and guidelines.

Emma Luk: A selection of my data science work

Appendix: Emma Luk (Mobile: 07974 522 805):

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

Deep learning is a subset of machine learning in artificial intelligence (AI) that is capable of learning from data.

Figure 1 shows Predicted Stock Prices (red) and Actual Stock Prices (blue)

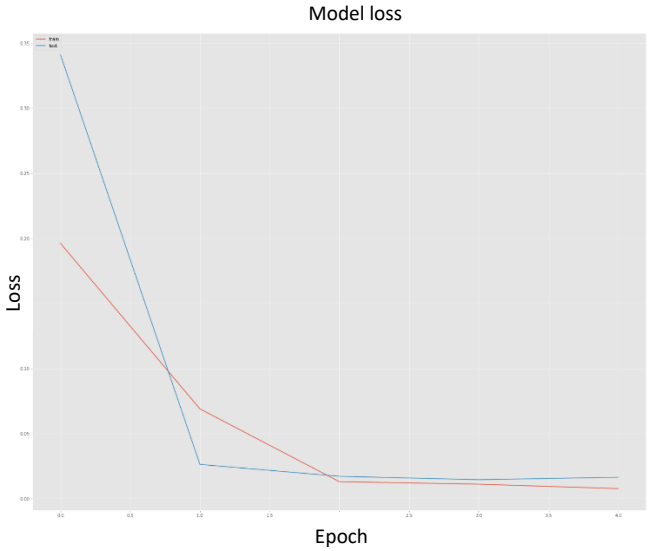


Figure 2 shows that a model is correct or just right training error (red line) slightly lower than test error (blue line).

PS: In deep learning, a loss function that quantifies the badness of our model, a model that is underfit will have high training and high testing error while an overfit model will have extremely low training error but a high testing error.

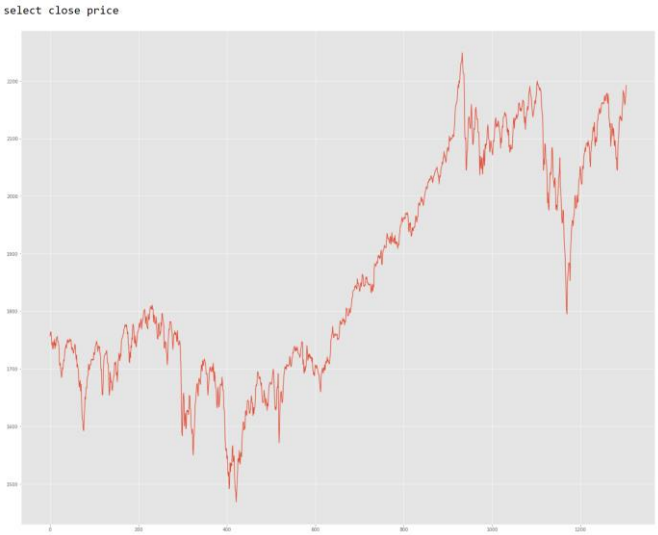
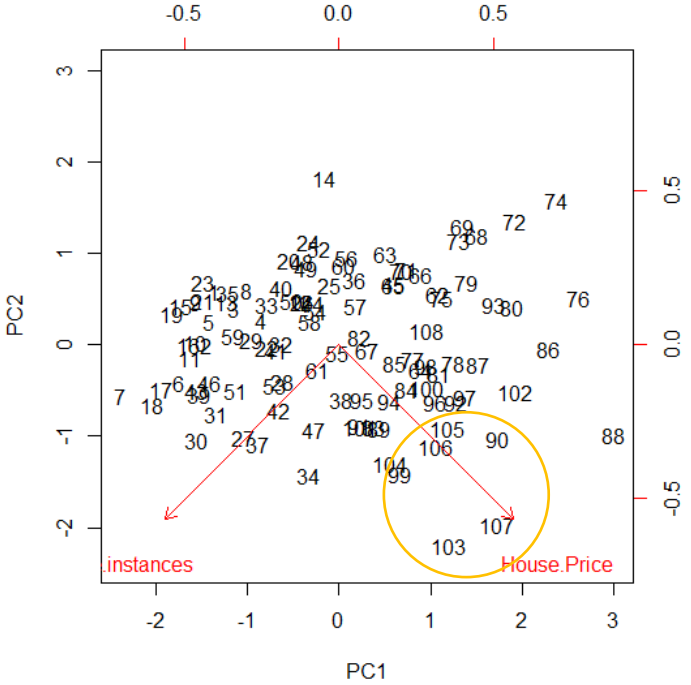


Figure 3: Plot of 'Close' for Global Equity Income sector price history

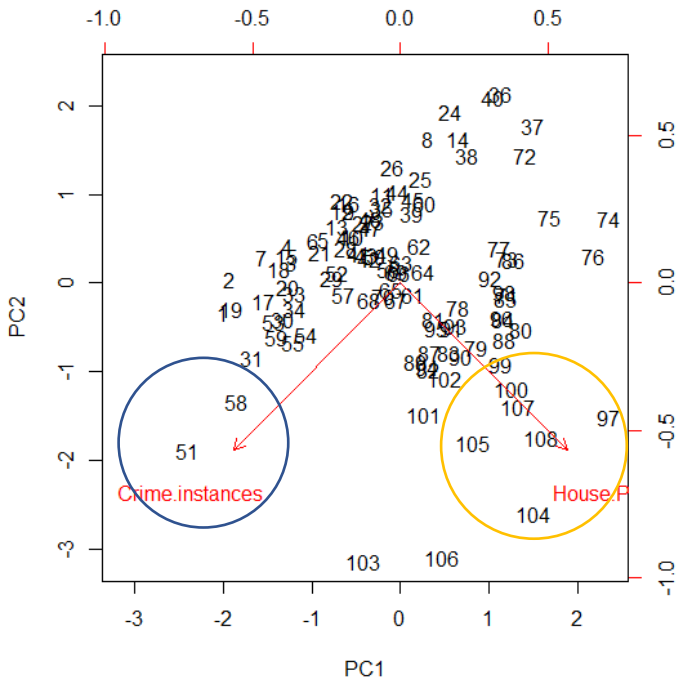
1B: Analysing House Prices and Crime in R

Unsupervised learning: Principal Component Analysis (PCA):

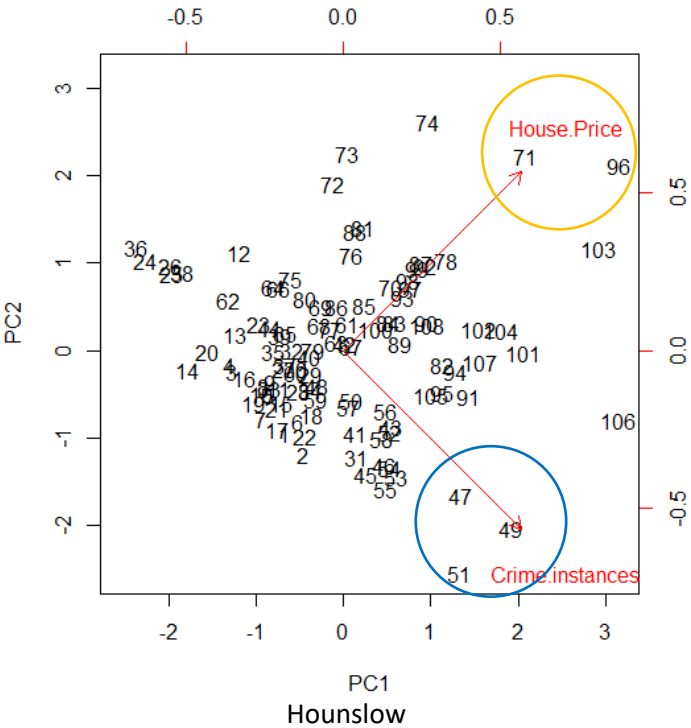
- The aim was to investigate how current house prices were affected by recent crime levels in London Boroughs
- Performed data cleaning, transformation, manipulation and conducted Principal Component Analysis (PCA): this was the process by which compute principal components and used them for **better 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

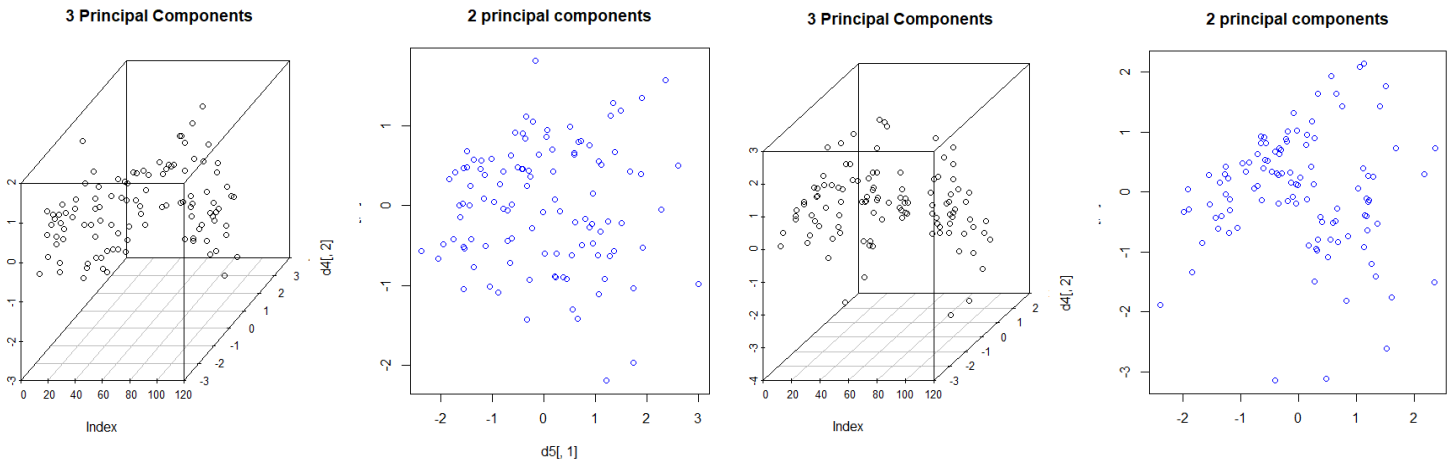
In Wandsworth:
House 51: £ 538,999.3
House 104: £ 973,938.4

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.

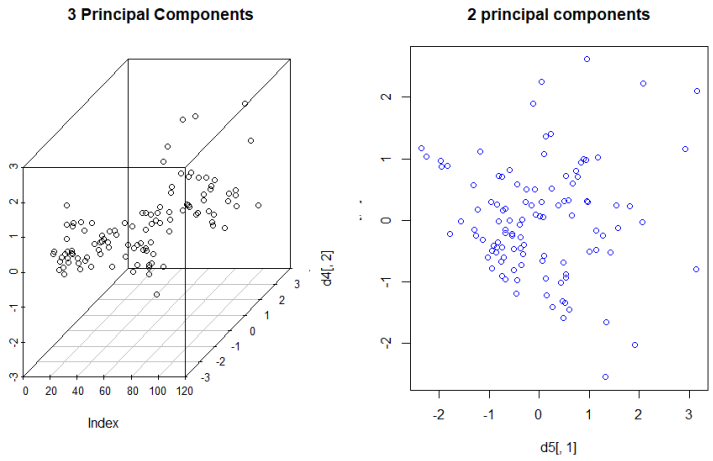
House 51 and 49 data points are near Crime point. They are less expensive. This is an indication that crime has some effect on house prices.

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



Merton

Wandsworth

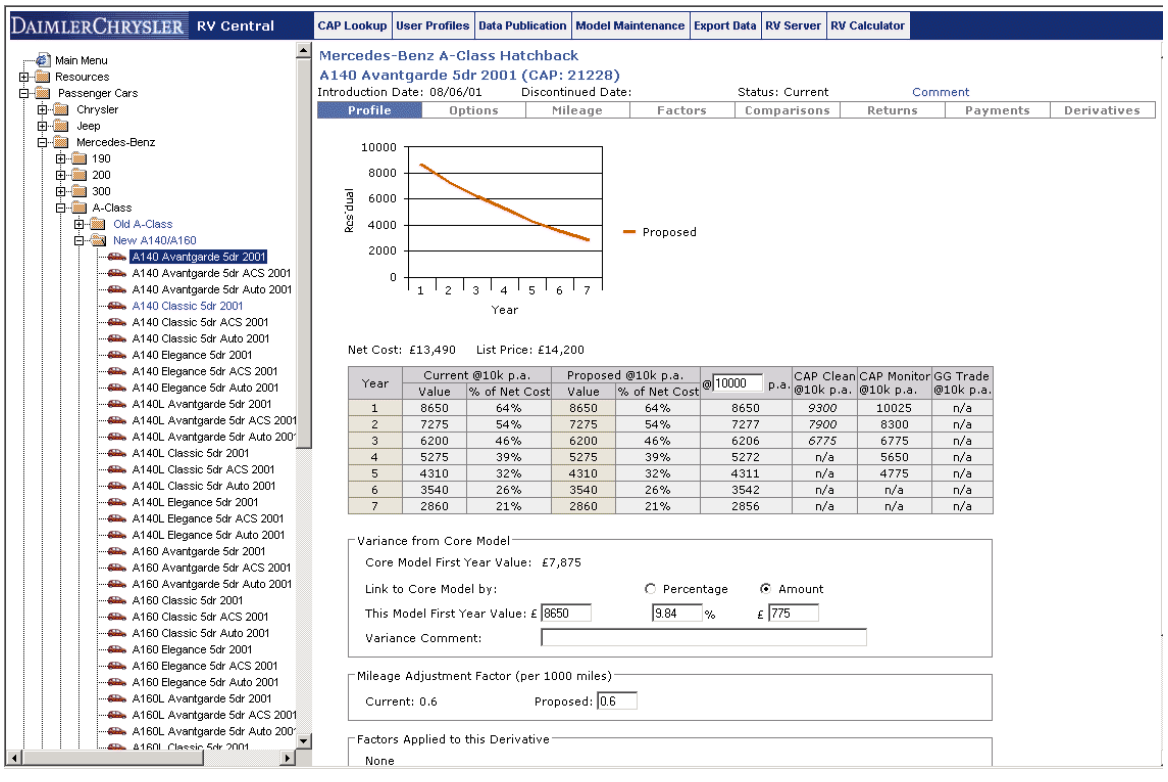


Hounslow

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

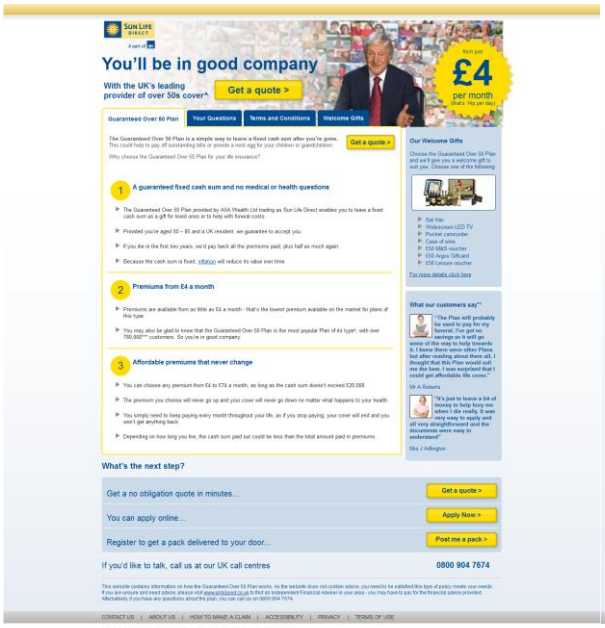
1C. An example from when I worked at Mercedes-Benz UK:

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.



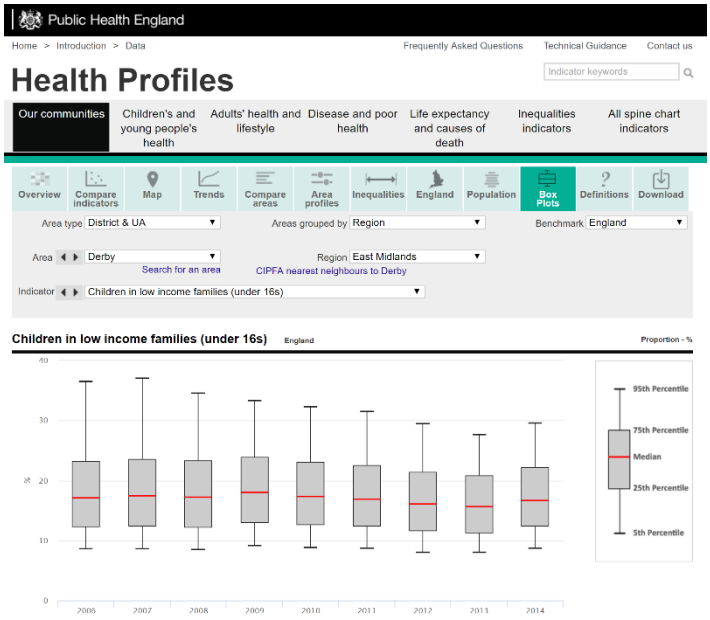
1D. Examples of A/B Testing and Multivariate Testing (MVT) from when I worked at AXA.

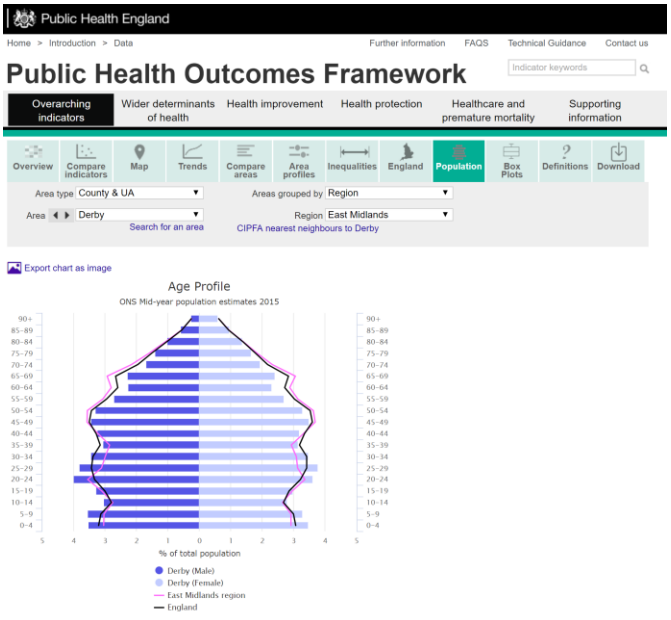
The two examples below depict different versions of the same webpage, which were used to provide insight to drive future strategies and identify business opportunities and problems.



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

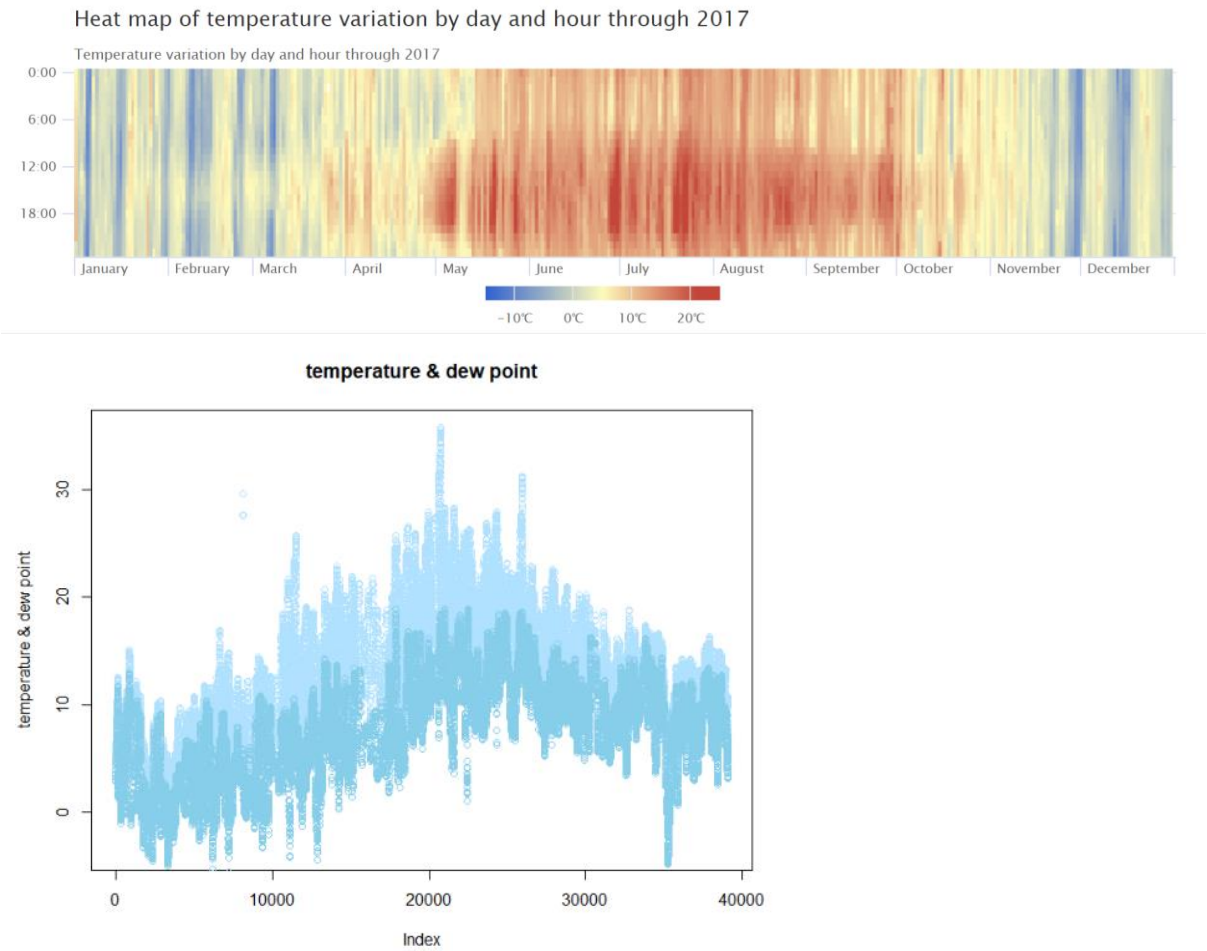


Figure 6: Scatter plots with two level
Added two level first level is dew point and second level is temperature in R

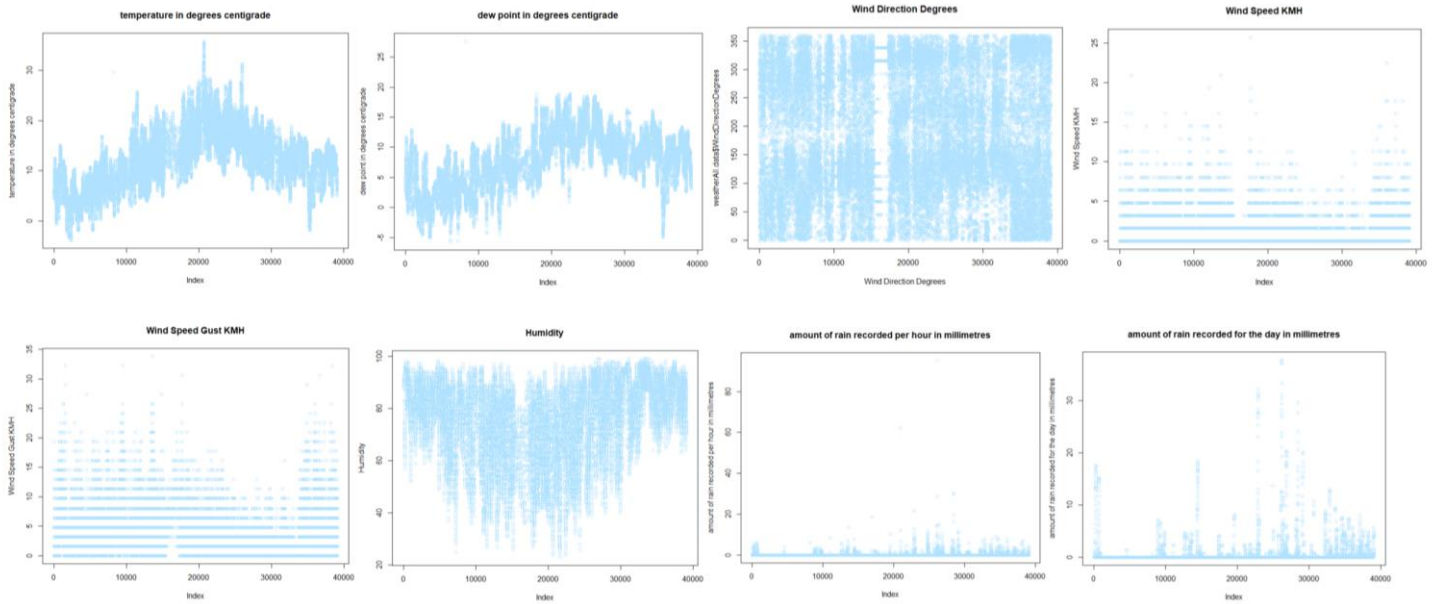
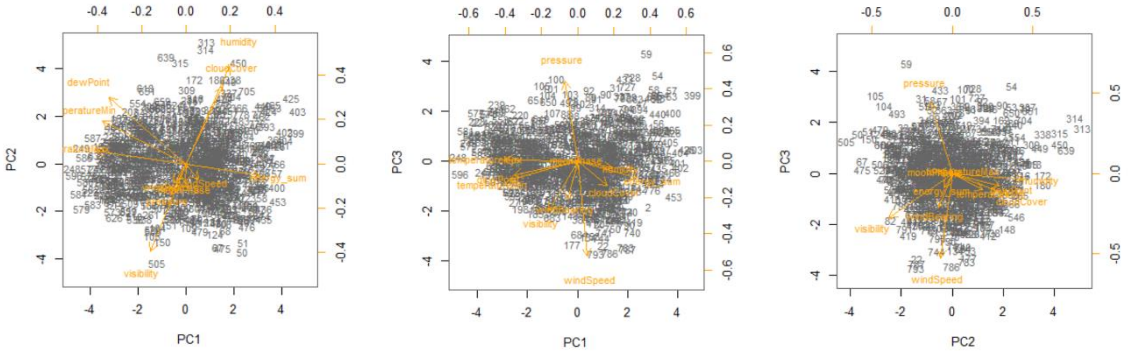


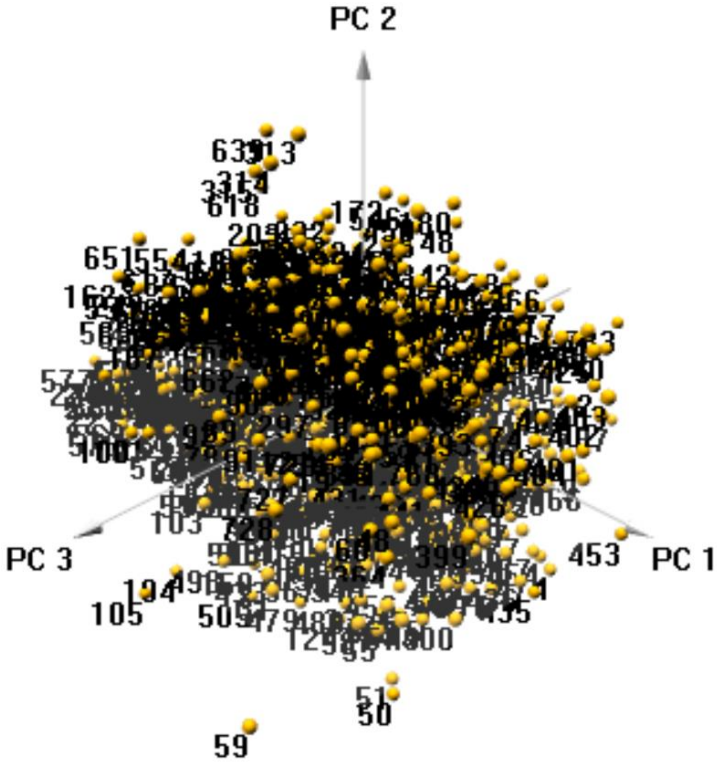
Figure 7: Scatter plots display the values of variables

1G: Explore relationships between weather conditions and energy consumption with R

- The goal was to draw a graph that shows how the samples are related (or not related) to each other.
- What was the relationships between weather conditions and energy consumption in London

Principal Component Analysis (PCA) Results with 2D & 3D graphs:





The project has over 11 features. PCA transformed variables into a new set of variables, which was a linear combination of the original variables.

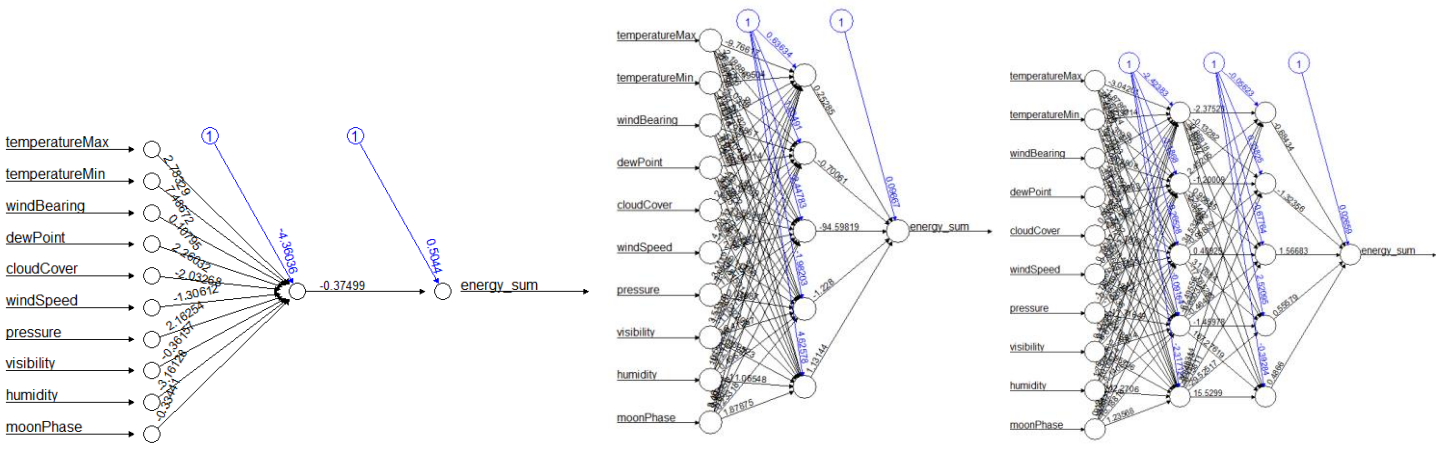
PCA is deterministic. So, the correct answer is guaranteed. It makes data plottable on a 2D graph.

PCA is a popular technique to transform a dataset onto a lower dimensional subspace for visualisation and further exploration. PC are Eigen-pairs. They describe the direction in the original feature space with the greatest variance.

House 59, 51 and 50: PCA sensitive to outliers and may cause wrong eigendirection.

Neural Networks

An artificial neural network was used. This kind of network has ten input layers, one output layer, and a number of hidden layers. The nodes in each layer are called neurons which perform non-linear calculations.



1H: 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:

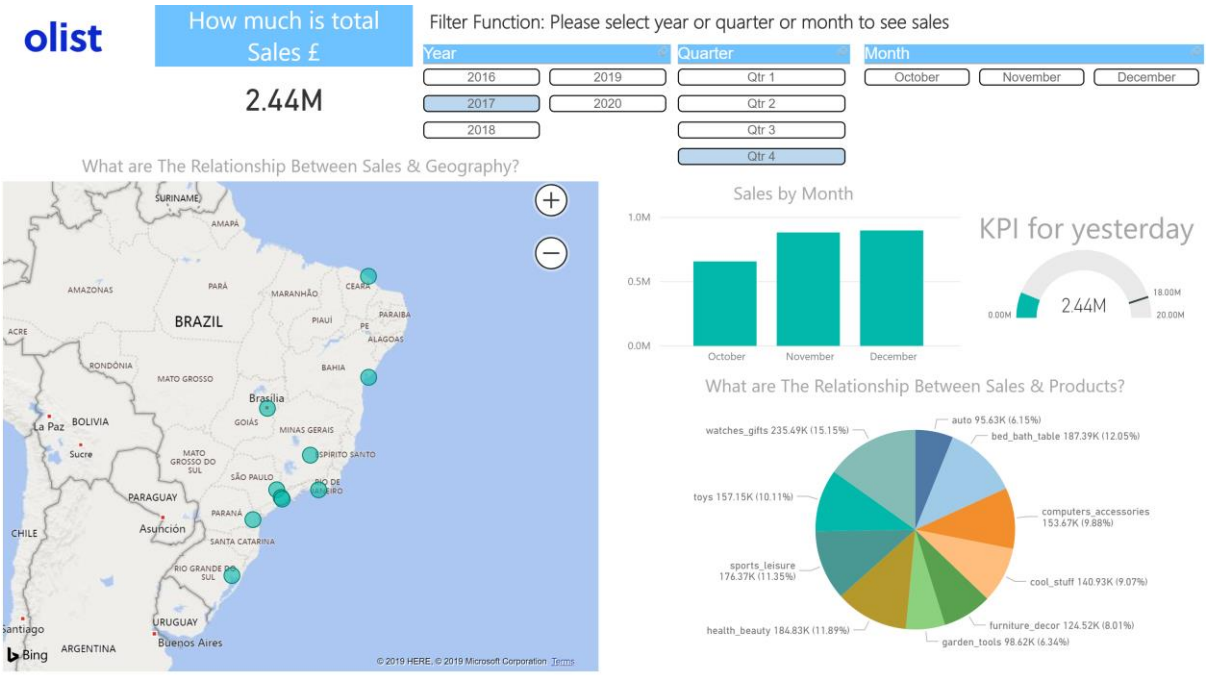
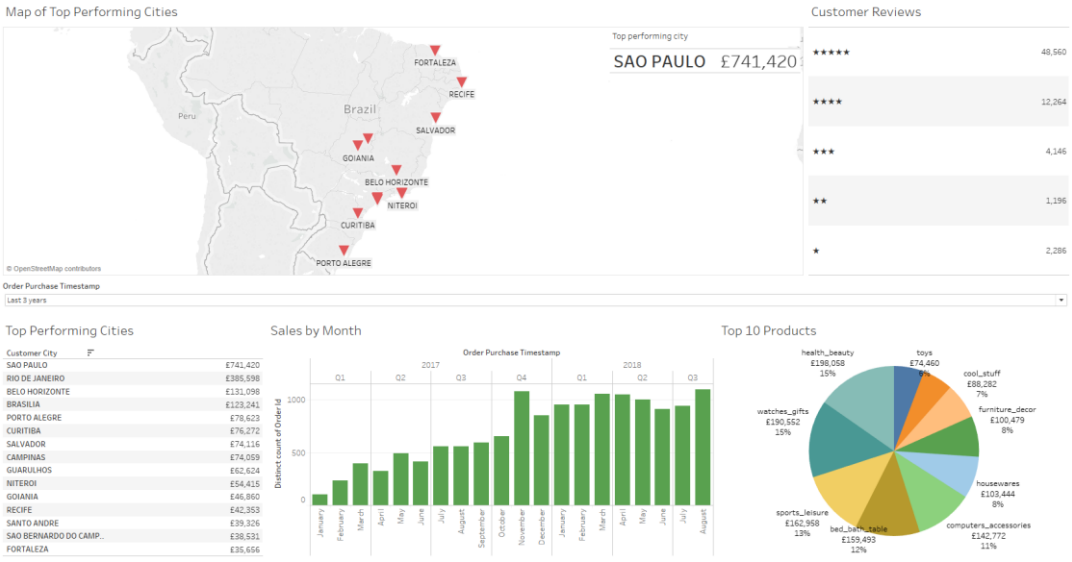


Tableau:



1I: Self-driving cars project: Control Unit

A collaboration between University College London (UCL) and MAL to build an autonomous car simulation using autonomous car point-cloud sensor data analysis, where applied data science and deep learning was used to build a control unit. A fuzzy time series model is implemented along with an LSTM-RNN deep neural network to produce the desired output.