

Emma Luk: A selection of my data science work

1. Maximising Sales through Optimal Marketing Budget Allocation: Leveraging Machine Learning for Personalised Marketing Insights

Introduction:

As an Online Marketing Webmaster at AXA and worked with multiple marketing channels, including television advertising, online display ads, and social media campaigns. The primary objective was to determine the optimal allocation of the marketing budget across these channels to maximise sales and deliver personalised experiences to customers.

To achieve data-driven decision-making, advanced techniques like A/B Testing and Multivariate Testing (MVT) were employed to continuously refine marketing strategies. These experiments allowed comparisons of different variations of campaigns, landing pages, and content to identify the most effective approaches for the target audience.

Additionally, the power of Machine Learning was harnessed to unveil valuable customer insights. Customer segmentation analysis revealed five distinct customer clusters based on behaviours, preferences, and needs. This understanding facilitated crafting hyper-personalised marketing messages tailored to each cluster's unique characteristics.

The primary objectives for this project include:

- **Marketing Budget Optimisation:** Analysing historical marketing data to identify channels with the highest returns and determining the most effective allocation of the marketing budget.
- **A/B Testing and MVT:** Conducting controlled experiments to test various marketing variables and campaign strategies, revealing combinations that drive optimal engagement and conversion rates.
- **Machine Learning Customer Clustering:** Leveraging Machine Learning algorithms to segment the customer base into five distinct clusters, enabling the creation of targeted marketing campaigns based on preferences and behaviours.
- **Personalised Marketing Approach:** Crafting personalised marketing strategies that cater to each segment's needs, fostering deeper connections with customers.

Through the integration of data-driven decision-making, innovative experimentation, and Machine Learning insights, the goal is to drive business growth, enhance customer experiences, and position AXA as a market leader in the insurance industry.

Throughout this journey, actionable insights will be showcased through visualisations, dashboards, or reports. Effective communication of results to stakeholders empowers AXA with the knowledge to make informed decisions and embark on a journey towards a more personalised and impactful marketing approach.

1A 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.

Achieved significant accomplishments during my tenure at AXA, including:

- Devised and implemented effective tracking metrics to monitor online customer behaviours.
- Built models to transform data points into actionable business insights.
- Utilised machine learning techniques to accelerate testing processes and enhance recommendation engines.
- Successfully contributed to a 30% increase in incremental revenue.



Key Points:

- A/B testing compares two versions (A and B) to determine which one performs better.
- The conversion rates for version A and version B were 0.1230 and 0.1360 respectively.
- The Z-score measures the deviation of the observed difference from the expected difference under the null hypothesis.
- The p-value represents the probability of observing a difference as extreme as the one observed, assuming no true difference between versions.
- The calculated Z-score of 0.8659 and p-value of 0.2681 indicated that the difference in conversion rates was not statistically significant.
- This means that the observed difference could be due to random chance, and it is not conclusive that one version performs better than the other.

1B Market Mixing Models project:

```
1 import pandas as pd
2 from sklearn.model_selection import train_test_split
3 from sklearn.linear_model import LinearRegression
4 from sklearn.metrics import r2_score
5
6 # Load the dataset
7 data = pd.read_csv('marketing_data.csv')
8
9 # Select the features (independent variables) and target (dependent variable)
10 X = data[['television_spending', 'online_spending', 'social_media_spending']]
11 y = data['sales_revenue']
12
13 # Split the data into training and testing sets (80% training, 20% testing)
14 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
15
16 # Create and train the Linear Regression model
17 model = LinearRegression()
18 model.fit(X_train, y_train)
19
20 # Make predictions on the test set
21 y_pred = model.predict(X_test)
22
23 # Evaluate the model's performance using R-squared (coefficient of determination)
24 r_squared = r2_score(y_test, y_pred)
25 print("R-squared:", r_squared)
26
27 # Print the coefficients of the model
28 coefficients = pd.DataFrame({'Feature': X.columns, 'Coefficient': model.coef_})
29 print(coefficients)
30
```

R-squared: 0.9007431826905605

	Feature	Coefficient
0	television_spending	4.845151
1	online_spending	3.107265
2	social_media_spending	1.815924

1. R-squared (R^2):

- R-squared is a statistical metric that measures how well the regression model fits the data. It indicates the proportion of the variance in the dependent variable (sales revenue) that can be explained by the independent variables ('television_spending', 'online_spending', and 'social_media_spending').
- The R-squared value is approximately 0.901, suggesting that about 90.1% of the variability in sales revenue is explained by the independent variables in the model. A higher R-squared value indicates that the model's predictions are closer to the actual values.

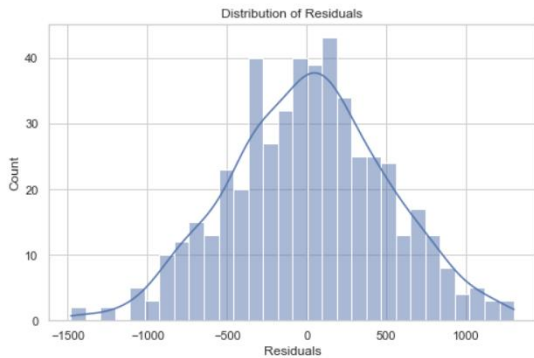
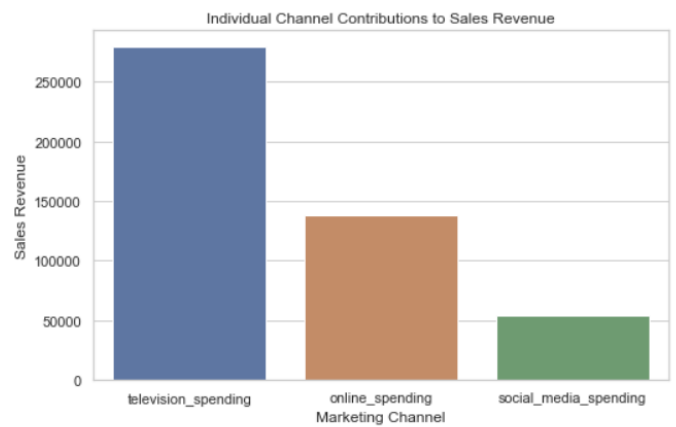
2. Coefficients:

- The coefficients represent the slope or effect of each independent variable on the dependent variable when holding other variables constant.
- For example, an increase of one unit in 'television_spending' is associated with an expected increase of approximately 4.85 units in sales revenue.
- Similarly, an increase of one unit in 'online_spending' is associated with an expected increase of around 3.11 units in sales revenue.
- Additionally, an increase of one unit in 'social_media_spending' is associated with an expected increase of about 1.82 units in sales revenue.

The coefficient values show the direction and magnitude of the relationships between the independent variables and the dependent variable. Positive coefficients indicate that an increase in the respective independent variable leads to an increase in the dependent variable (sales revenue), while negative coefficients suggest a decrease in the dependent variable with an increase in the independent variable.

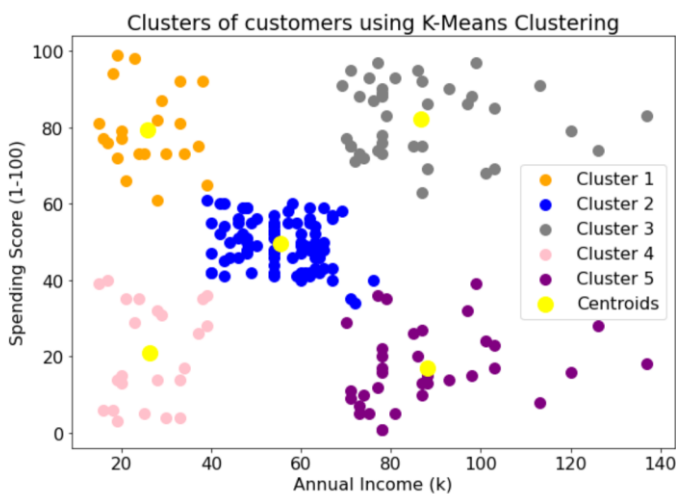
In summary, the R-squared value of approximately 0.901 indicates a reasonably good fit of the model to the data, and the coefficients provide insight into the contributions of each marketing channel (television, online, and social media spending) to the sales revenue in the Market Mixing Model project.

An example of visualising the MMM analysis results. It creates scatter plots to compare actual sales revenue with predicted sales, a distribution plot to show the residuals' distribution, and a bar plot to visualise the individual contributions of each marketing channel to sales revenue.



	sales_revenue	television_spending	online_spending	social_media_spending
count	500.000000	500.000000	500.000000	500.000000
mean	4885.660073	558.358000	275.740000	108.598000
std	1434.653693	263.241421	128.150214	52.234867
min	1189.707963	101.000000	50.000000	20.000000
25%	3824.813023	328.250000	170.500000	63.750000
50%	4830.691333	567.500000	272.000000	109.000000
75%	6056.205191	784.250000	390.250000	153.000000
max	8382.015677	999.000000	499.000000	199.000000

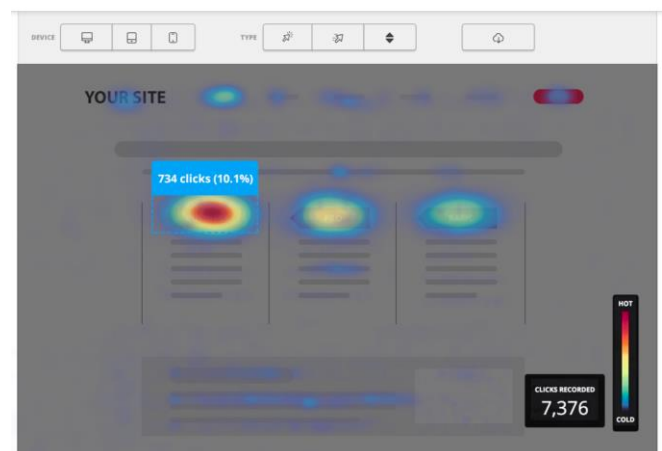
1C Leveraging Machine Learning for Customer Understanding and Personalised Marketing: Unveiling Five Distinct Customer Clusters



Just looking at the graph tells us about the five different types of customers

- I. Cluster 1: Low income, High spenders
- II. Cluster 2: Average income, Average expenditure
- III. Cluster 3: High income, High spenders
- IV. Cluster 4: Low income, Low Spenders
- V. Cluster 5: High income, Low spenders

Machine learning is used to understand customers, drive personalisation, streamline processes and create convenient and memorable customer experiences.



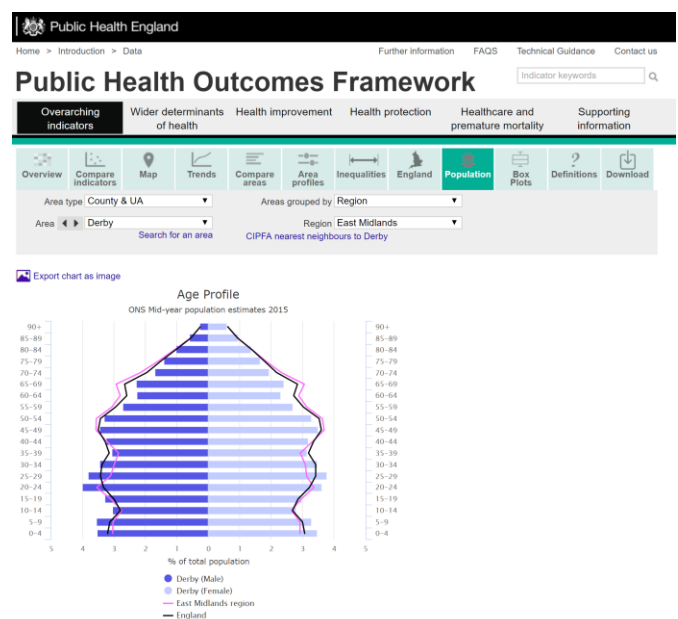
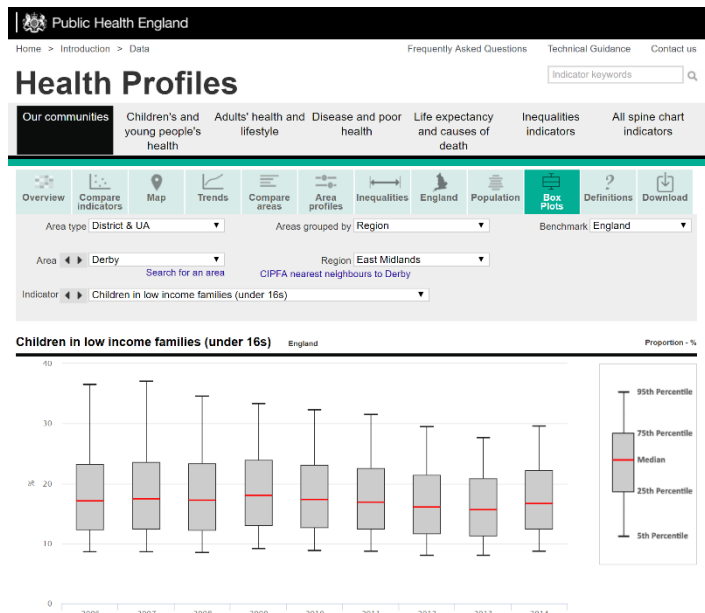
Rather than relying on a marketer's intuition to separate customers into groups for campaigns, data scientists can use clustering and classification algorithms to group customers into personas based on specific variations among them with similar needs and behaviours. These personas account for customer differences across multiple dimensions such as demographics, browsing behaviour, and affinity. Connecting these traits to patterns of purchasing behaviour allows data-savvy companies to roll out highly personalised marketing campaigns that are more effective at boosting sales than generalised campaigns.

Achieved significant accomplishments during my tenure at AXA, including:

- Developed a customer segmentation strategy using machine learning techniques to identify five distinct customer clusters based on income and spending behaviour.
- Implemented clustering and classification algorithms to group customers into personas, allowing for personalised marketing campaigns based on specific customer needs and behaviours.
- Utilised multiple dimensions such as demographics, browsing behaviour, and affinity to create comprehensive customer personas.
- Connected customer traits to patterns of purchasing behaviour, resulting in highly personalised and effective marketing campaigns.
- Contributed to streamlining processes and enhancing customer experiences by leveraging machine learning insights.
- Shifted marketing strategies from intuition-based decisions to data-driven approaches, resulting in improved sales and customer engagement.
- Led efforts to understand customers through machine learning, driving personalisation and creating convenient and memorable customer experiences.

2. 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.



- **Successfully contributed to data visualisation projects while working on the 'Fingertips' website for Public Health England (Government Agency), including:**
 - Created informative boxplots showcasing the percentage of children in low-income families in the East Midlands region from 2006 to 2014, effectively conveying the distribution and trends over time.
 - Developed a visually compelling bar graph with a negative stack to represent the proportion of males and females in different age groups within the East Midlands region, providing insights into demographic patterns and gender distribution.
 - Leveraged data visualisation techniques to present complex information in a clear and concise manner, facilitating understanding and communication of key findings to stakeholders.
 - Ensured visualisations adhered to best practices in data visualisation, employing appropriate design choices to enhance readability and interpretability.
 - Played a crucial role in making data-driven insights accessible and actionable, empowering decision-makers to utilise the visualisations to drive strategies and policies.
 - Collaborated with cross-functional teams to gather requirements, identify key metrics, and deliver impactful visualisations that supported evidence-based decision-making processes.

3: 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 cities, customer reviews and sales by Month.

Demonstrated expertise in data visualisation using Microsoft Power BI and Tableau, including:

- Created an Executive Dashboard to track and report on business metrics and key performance indicators (KPIs).
- Developed visually appealing and informative visualisations to provide actionable insights to stakeholders.
- Incorporated various data sources to create a comprehensive view of business performance, including top ten products, top-performing cities, customer reviews, and sales by month.
- Utilised interactive features and drill-down capabilities to enable users to explore data and gain deeper insights.
- Collaborated with cross-functional teams to gather requirements, design visually impactful dashboards, and deliver timely and relevant information to decision-makers.
- Ensured the Executive Dashboard adhered to best practices in data visualisation, including clear data storytelling, effective use of colours and layout, and user-friendly navigation.

Microsoft Power BI:

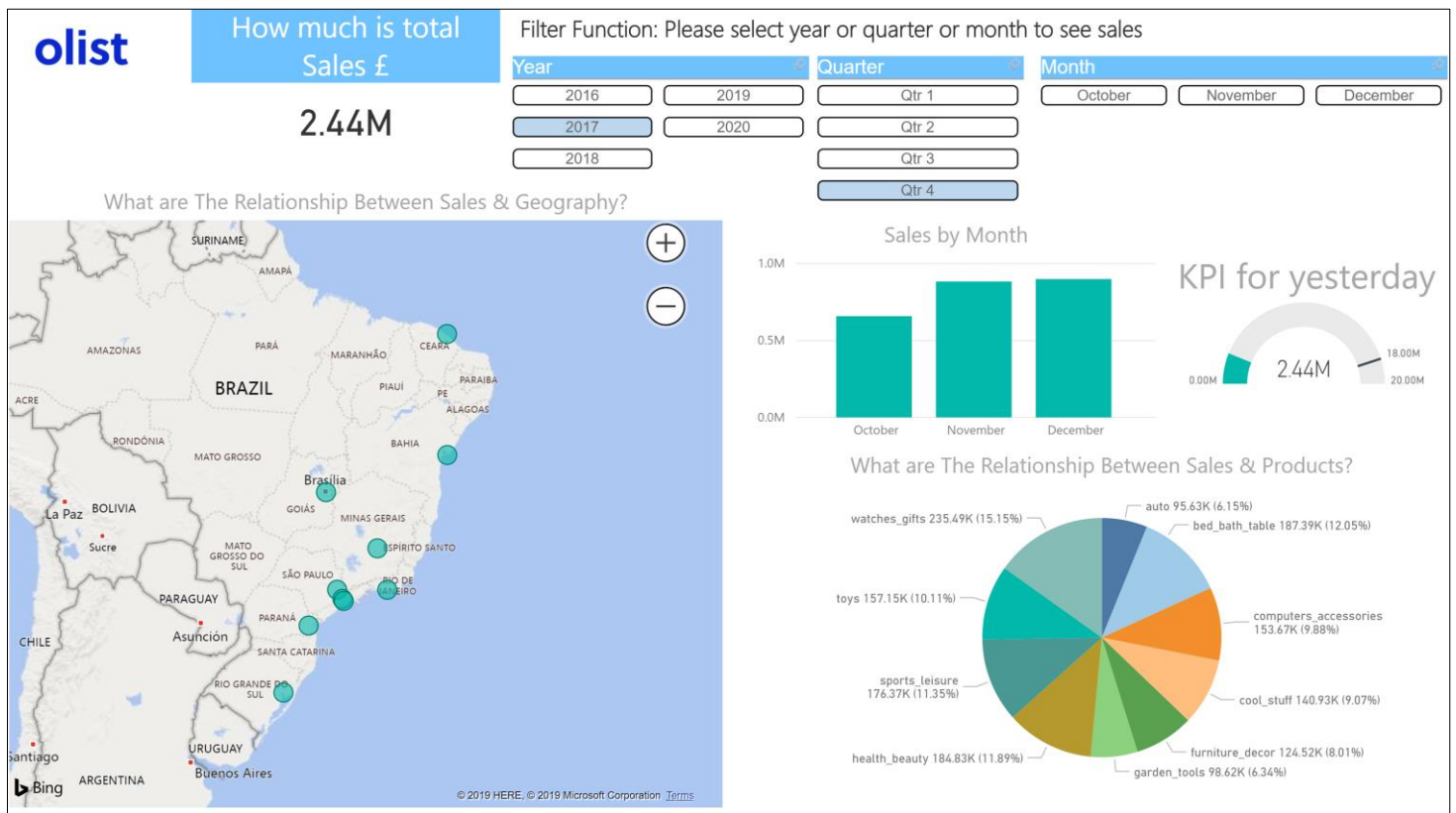
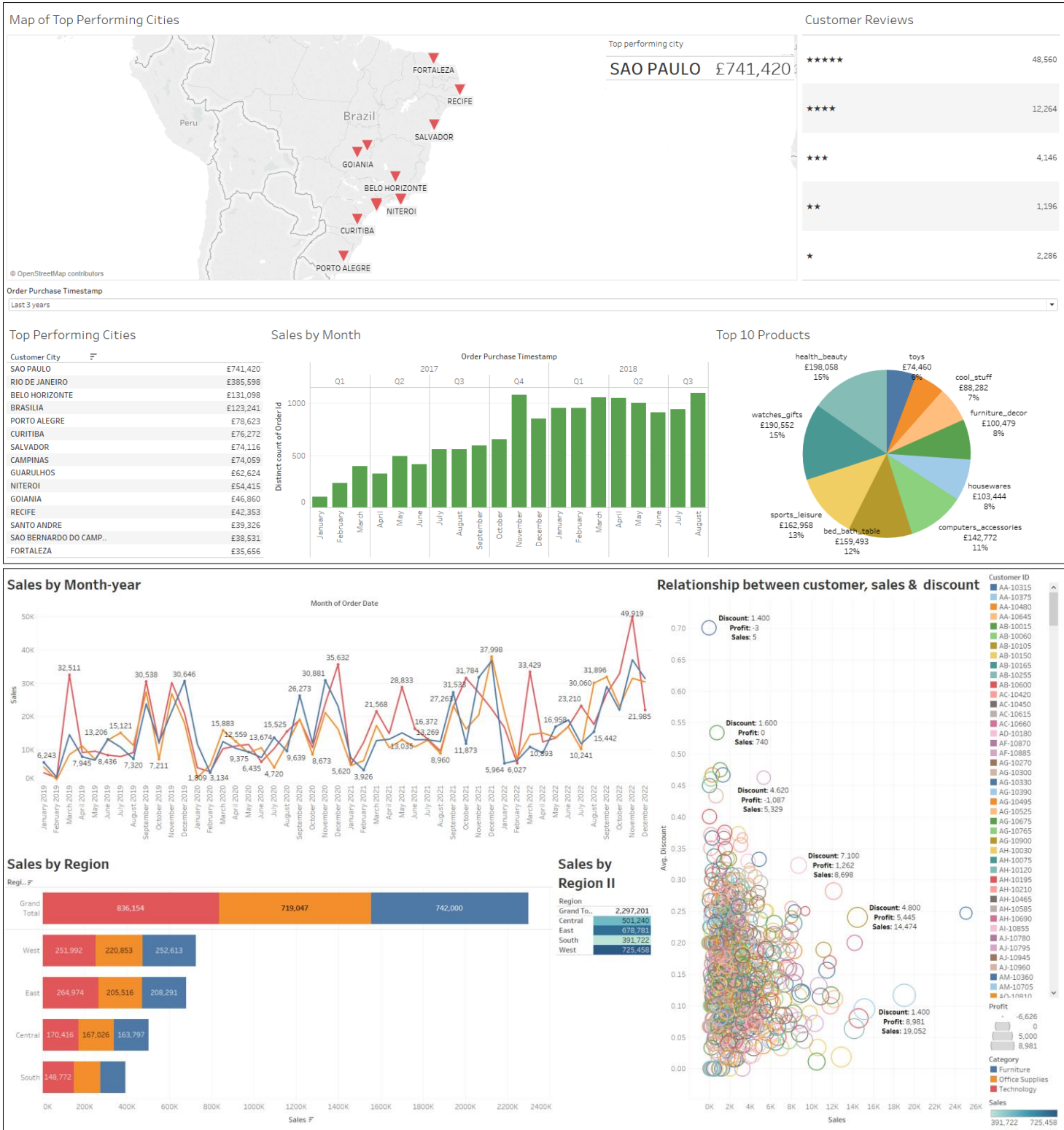


Tableau:



Sales by Month-year

Month of Order Date

January 2019

February 2019

March 2019

April 2019

May 2019

June 2019

July 2019

August 2019

September 2019

October 2019

November 2019

December 2019

January 2020

February 2020

March 2020

April 2020

May 2020

June 2020

July 2020

August 2020

September 2020

October 2020

November 2020

December 2020

January 2021

February 2021

March 2021

April 2021

May 2021

June 2021

July 2021

August 2021

September 2021

October 2021

November 2021

December 2021

January 2022

February 2022

March 2022

April 2022

May 2022

June 2022

July 2022

August 2022

September 2022

October 2022

November 2022

December 2022

Sales by Region

Region, £

Grand Total

836,154

719,047

742,000

West

251,992

220,853

252,613

East

264,974

205,516

208,291

Central

170,416

167,026

163,797

South

149,772

Sales by Region II

Region

Grand To..

Central

East

South

West

2,297,201

501,240

678,781

391,722

725,458

Relationship between customer, sales & discount

Customer ID

AA-10315

AA-10375

AA-10480

AB-10015

AB-10060

AB-10105

AB-10150

AB-10165

AB-10255

AB-10600

AC-10420

AC-10450

AC-10615

AC-10660

AF-10870

AF-10885

AG-10270

AG-10300

AG-10330

AG-10390

AG-10495

AG-10525

AG-10675

AG-10765

AG-10900

AH-10030

AH-10075

AH-10120

AH-10195

AH-10210

AH-10465

AH-10585

AH-10690

AI-10855

AJ-10780

AJ-10795

AJ-10945

AM-10360

AM-10705

an-10610

Profit

-5,626

0

5,000

8,981

Category

Furniture

Office Supplies

Technology

Sales

391,722

725,458

4. Data Scientist: Driving Data-Driven Insights for Strategic Marketing Success using Google BigQuery ([Appendix 4](#))

- **Successful Business Case Analysis ([Appendix 4](#)):** Conducted an in-depth analysis of 100k orders at Target in Brazil, identifying key relationships between customer, order, seller, payment, review, product, and geolocation datasets. This analysis provided valuable insights for customer segmentation and sentiment analysis.
- **Data Exploration and Understanding:** Proficiently imported and analysed complex datasets, employing exploratory analysis techniques to identify dataset structures, characteristics, and relationships. Demonstrated expertise in handling large datasets and extracting actionable insights.
- **Effective Data Integration:** Implemented efficient data integration strategies to establish relationships between diverse datasets, such as customers and orders, sellers and orders, payments and orders, reviews and orders, products and order items, and geolocation and customers/sellers. Ensured seamless data flow and enabled comprehensive analysis.
- **Economic Impact Analysis:** Conducted an in-depth analysis of the money movement in the e-commerce industry, focusing on order prices, freight, and other factors. Calculated the percentage increase in the cost of orders from 2017 to 2018, specifically for the months between January and August. Provided valuable insights into the changing economic landscape of the industry.
- **Technical Expertise:** Demonstrated proficiency in utilising tools like BigQuery and coding functions like EXTRACT, SUM, and conditional aggregation to perform complex calculations and generate meaningful results. Leveraged advanced analytical techniques to drive data-informed decision making.
- **Strategic Insights:** Translated complex data analysis into actionable insights, contributing to strategic marketing initiatives. Collaborated with cross-functional teams to leverage data-driven approaches and optimise marketing campaigns.
- **Communication and Presentation:** Effectively communicated analysis findings and insights to both technical and non-technical stakeholders. Presented complex analytical concepts in a clear and concise manner, facilitating informed decision making at the organisational level.