

Capstone Project

**Customer Insights and
Analysis**

By

Yogita Mane

15th April '23

Table of Contents

Problem Statement	3
Industry/ domain	3
Stakeholders.....	3
Business question.....	3
Data Question	3
Data	3
Data Science Process.....	5
Data Analysis	5
Feature Engineering	13
Modelling	16
Outcome	19
Implementation	19
Data answer	19
Business answer	19
Response to Stakeholders	19
End-to-end solution	19
References.....	21

Problem Statement

The Telecommunication Industry has a diverse user base, making it challenging for companies to identify patterns and trends to meet unique user needs. Complex user journeys with multiple touchpoints, channels, and devices can make it difficult to track and optimize user experience. Lack of reliable and complete data on user behaviour and engagement is a significant obstacle, despite collecting vast amounts of data. Additionally, strict data privacy and security regulations limit the amount of data available to companies, making it challenging to gain a comprehensive understanding of user behaviour and engagement.

Industry/ domain

This is a telecommunications company located in New Zealand that offers a range of services to its customers. These services include mobile and fixed-line telecommunications, internet and broadband services, cloud computing, digital media, and IT services. In addition, the company also sells a variety of devices, such as smartphones, tablets, and laptops.

The company is dedicated to staying ahead of the game and keeping up with its customers' ever-changing needs through innovation and investment in its network infrastructure and digital services. It takes customer support very seriously and offers a variety of channels for customers to seek assistance, including self-service options.

This telecommunications company is known for its dynamic and innovative approach and plays a crucial role in connecting people and business not only in New Zealand but all around the world.

Stakeholders

The Marketing Lead introduced the dataset in question to gain insights into age & gender behaviour on their e-commerce platform, facing challenges such as lack of understanding target audience, ineffective messaging, poor user experience, and inadequate data analysis.

Business question

"What strategies can we use to improve the conversion rates of our e-commerce platform for different age groups and genders?" We want to understand their engagement patterns, so that we can tailor our offerings to meet their needs and increase their likelihood of making a purchase.

Data Question

"What data variables are needed to accurately analyse customer behaviour and identifying opportunities for optimisation".

This will involve collecting and analysing data on key metrics such as bounce rate, conversion rate and demographic information.

Data

Dataset submitted is a company data. This comprehensive dataset offers insights into user engagement patterns and e-commerce performance across different age groups and genders.

Data is sourced from Google Analytics, this allows business to track and report website traffic as a platform inside the Google Marketing Platform. In future, this dataset can be made available through Marketing Team.

This dataset contains various variables that represent specific aspects of user engagement and e-commerce performance on a website.

By carefully analysing this data, business can uncover valuable trends and opportunities to optimize their website content, improve user experience, and refine marketing campaigns to increase e-commerce conversion rates and generate more revenue.

Data Science Process

Data Analysis

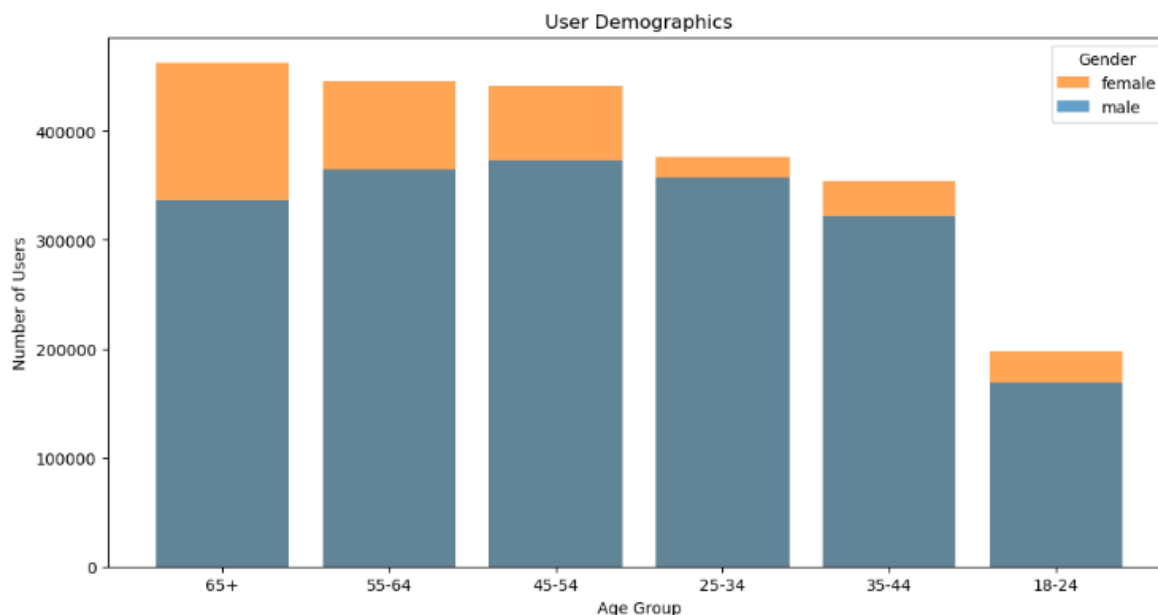


Figure 1: User Demographic Bar Chart

Analysing user demographics provides valuable insights into the different age and gender groups that engage with a business's website or online platform. This understanding can help business make informed decisions about marketing, product development, and customer service. Here is what the demographic data means for a business:

Age Distribution: The data indicates that most users fall into the age groups of 65+, 55-64, and 45-54, with a relatively even distribution across these groups. This suggests that the business's products or services may appeal more to an older audience. To target these age groups more effectively, marketing efforts can be tailored accordingly.

Gender Distribution: Female users are more prevalent in most age groups, particularly in the 65+, 55-64, and 45-54 age groups. This implies that the business's offerings may cater more to the interests and needs of women. To better cater to this demographic, it may be beneficial for the business to develop more female-centric marketing campaigns or adjust their product offerings.

Bounce rate

According to the data, it appears that there is a slightly higher bounce rate among males in certain age groups, specifically those aged 25-34 and 35-44. In fact, the highest bounce rate for males is observed in the 25-34 age group at 69.67%, whereas the highest bounce rate for females is in the 65+ age group at 67.26%.

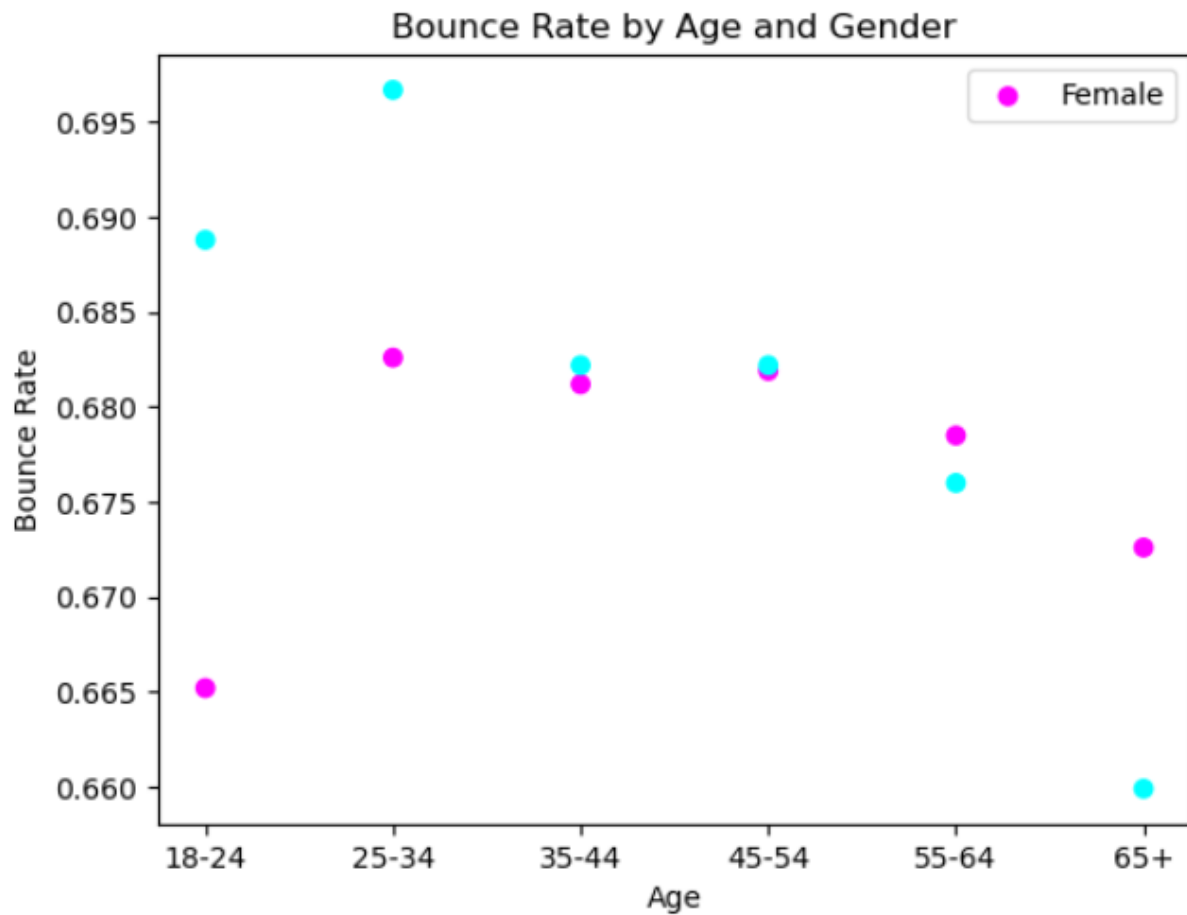


Figure 2: Scatter Plot showing Bounce rate

This data is indicative of a potential issue with the website's content or user experience, particularly for male users in these age groups. It may be worthwhile for the business to conduct further investigation into this matter to enhance user engagement and reduce bounce rates.

Page per session

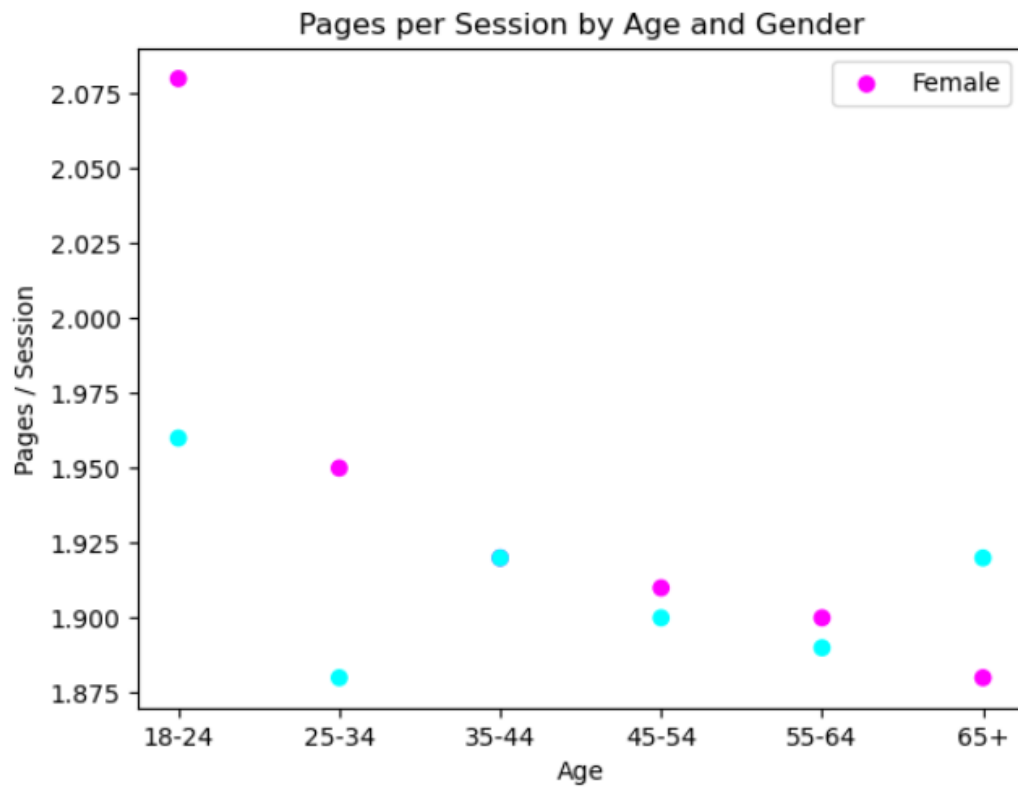


Figure 3: Scatter plot showing Pages per Session

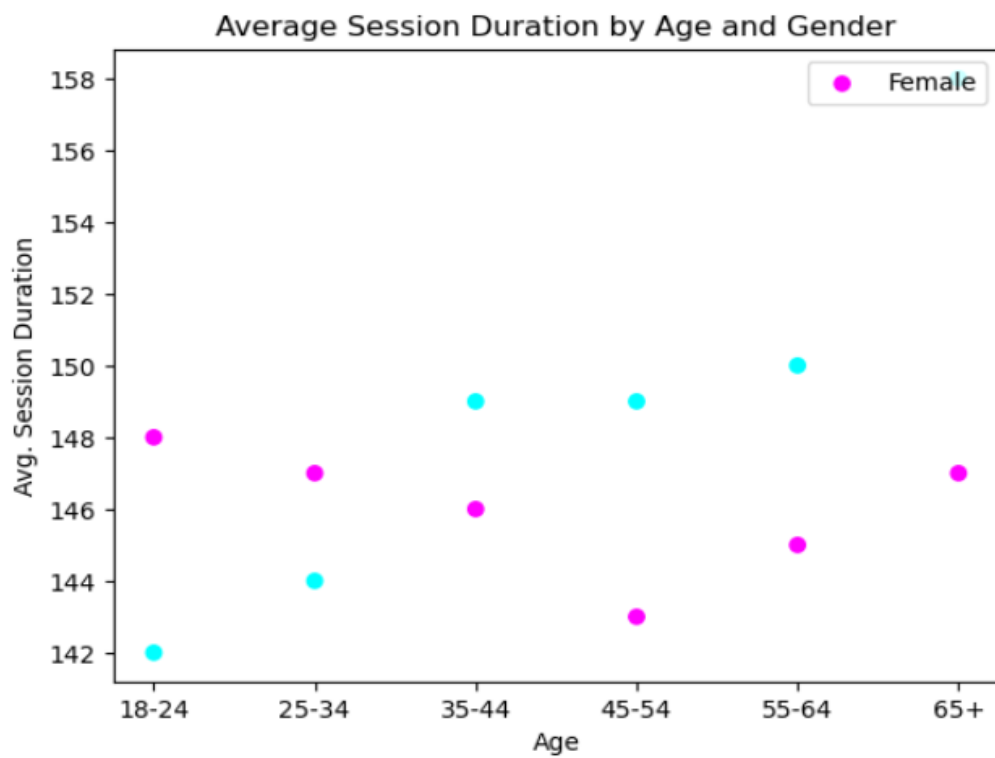


Figure 4: Scatter plot showing Average Session Duration

Based on our analysis, we have found that younger age groups, specifically those aged 18-24 and 25-34, tend to spend more time on the website and view more pages than older age groups. Additionally, we have observed that females within the 18-24 age group exhibit the highest pages per session and average session duration, indicating a high level of engagement with the website's content among this demographic. These insights suggest that tailoring content and user experience towards younger age groups and female audiences could lead to increased engagement and higher levels of satisfaction among our users.

Transactions, Revenue, and E-commerce conversion rate

When we analyse the data related to transactions, revenue, and e-commerce conversion rate, it's evident that the younger age groups, specifically those between 18-24 and 25-34 years, exhibit a higher number of transactions and revenue generated from the website.

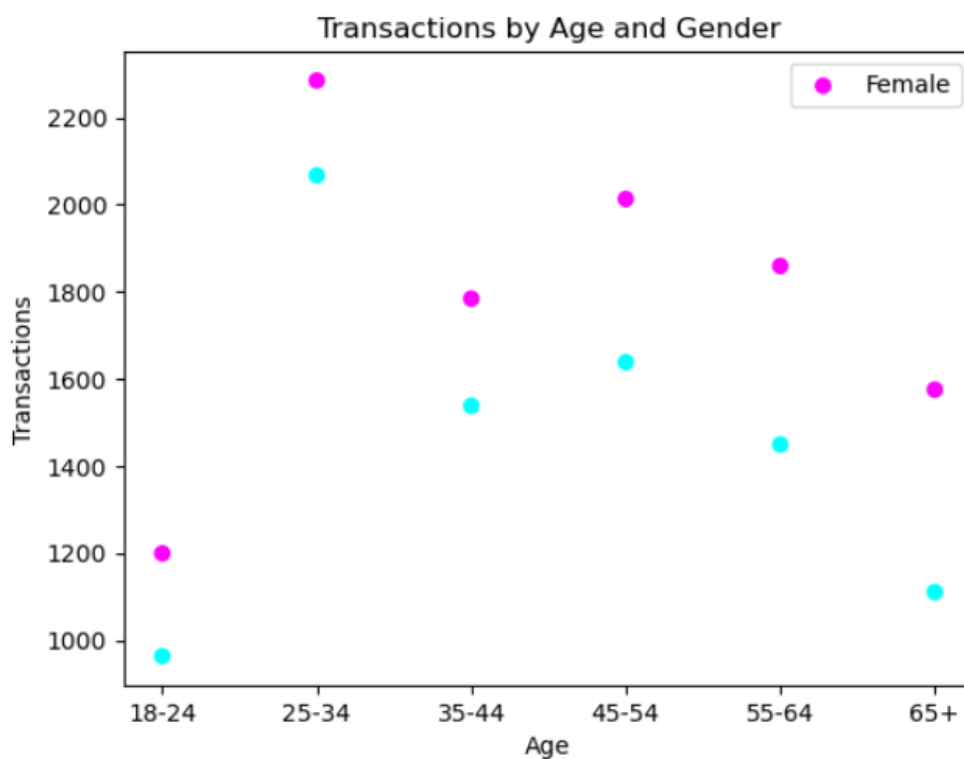


Figure 5: Scatter plot showing transactions

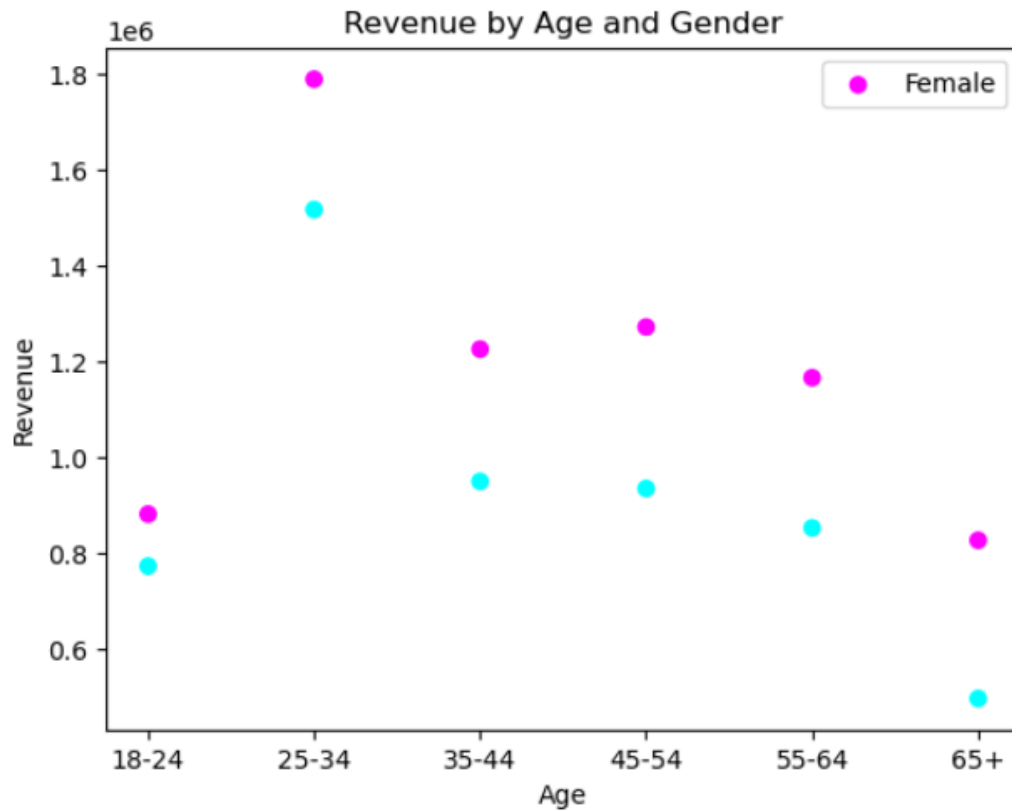


Figure 6: Scatter plot showing Revenue

The e-commerce conversion rate is also observed to be the highest in these age groups, implying that they are more likely to make a purchase from the website as compared to other age groups.

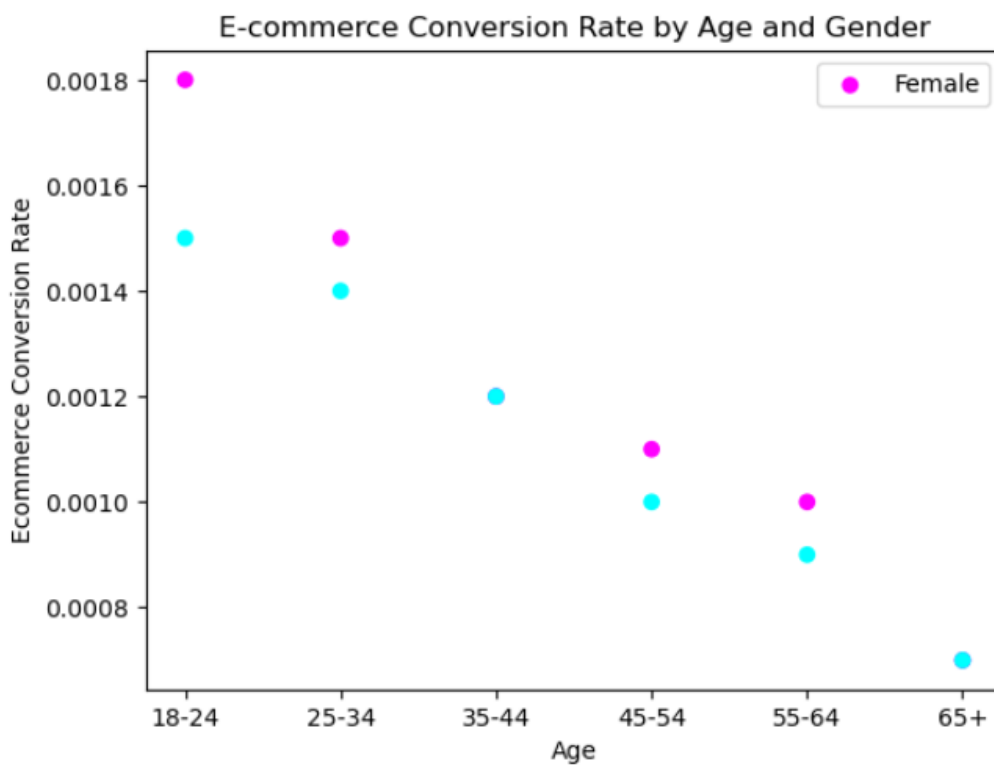


Figure 7: Scatter plot showing E-commerce conversion rate

Upon careful examination, we can draw the conclusion that younger age groups, especially females in the 18-24 age group, are deeply involved with the website and are more inclined towards making a purchase. This information can prove to be extremely valuable for business, allowing them to focus their marketing efforts on these demographics and enhance engagement and revenue. Additionally, companies can work towards improving the website's content and user experience for older age groups, thereby increasing engagement, and reducing the bounce rate.

Pages per session and Average Session Duration

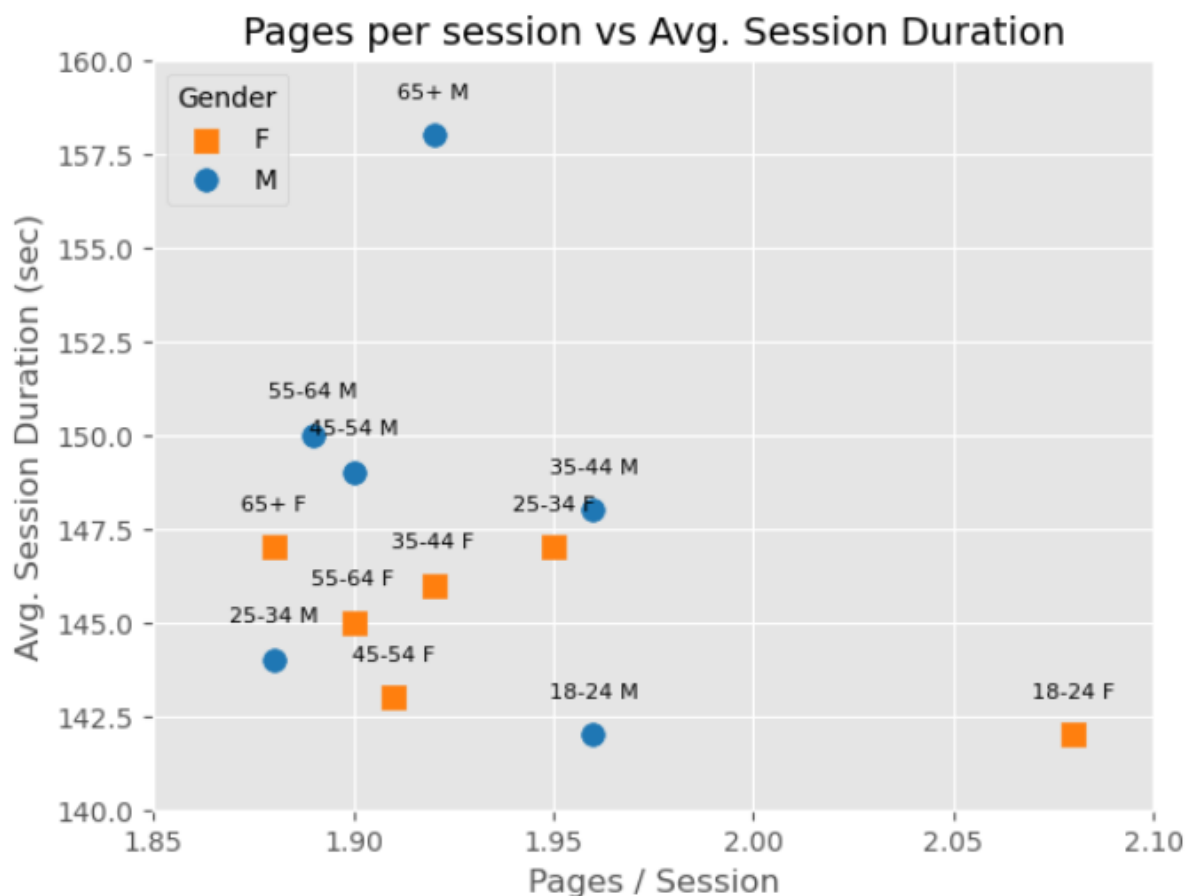


Figure 8: Scatter plot showing Pages per session vs Av. Session Duration

Pages/Session refers to the average number of pages viewed in a website session. In this dataset, Pages/Session range from 1.88-2.08, with the highest value for 18-24 female age group.

Avg. Session Duration is the average time spent on the website per session, with a range of 00:02:22-00:02:38, highest value for 65+ male age group.

E-commerce Performance

	Highest Transactions	Highest Revenue	Highest Conversion Rate
Gender	female	female	female
Age	25-34	25-34	18-24
Value	2285	\$1,788,988.56	0.18%

Here are the key takeaways from the data analysis:

The data indicates that female customers aged 25-34 are the most active and engaged in making purchases on our e-commerce platform. They also generate the highest revenue for the business, making them the most valuable customer segment.

Furthermore, females aged 18-24 have the highest e-commerce conversion rate, meaning that they are more likely to make a purchase on our platform than any other age group.

With these insights, we can focus our efforts on targeting these valuable customer segments through personalized promotions and recommendations. By improving their overall experience on our platform, we can increase customer loyalty and retention.

Finally, we can use this data to better understand the needs and preferences of our high-performing customer segments and develop new products or services that cater to their specific interests. By doing so, we can attract new customers and continue to grow our business.

Correlation analysis between user engagement metrics and e-commerce conversion rates

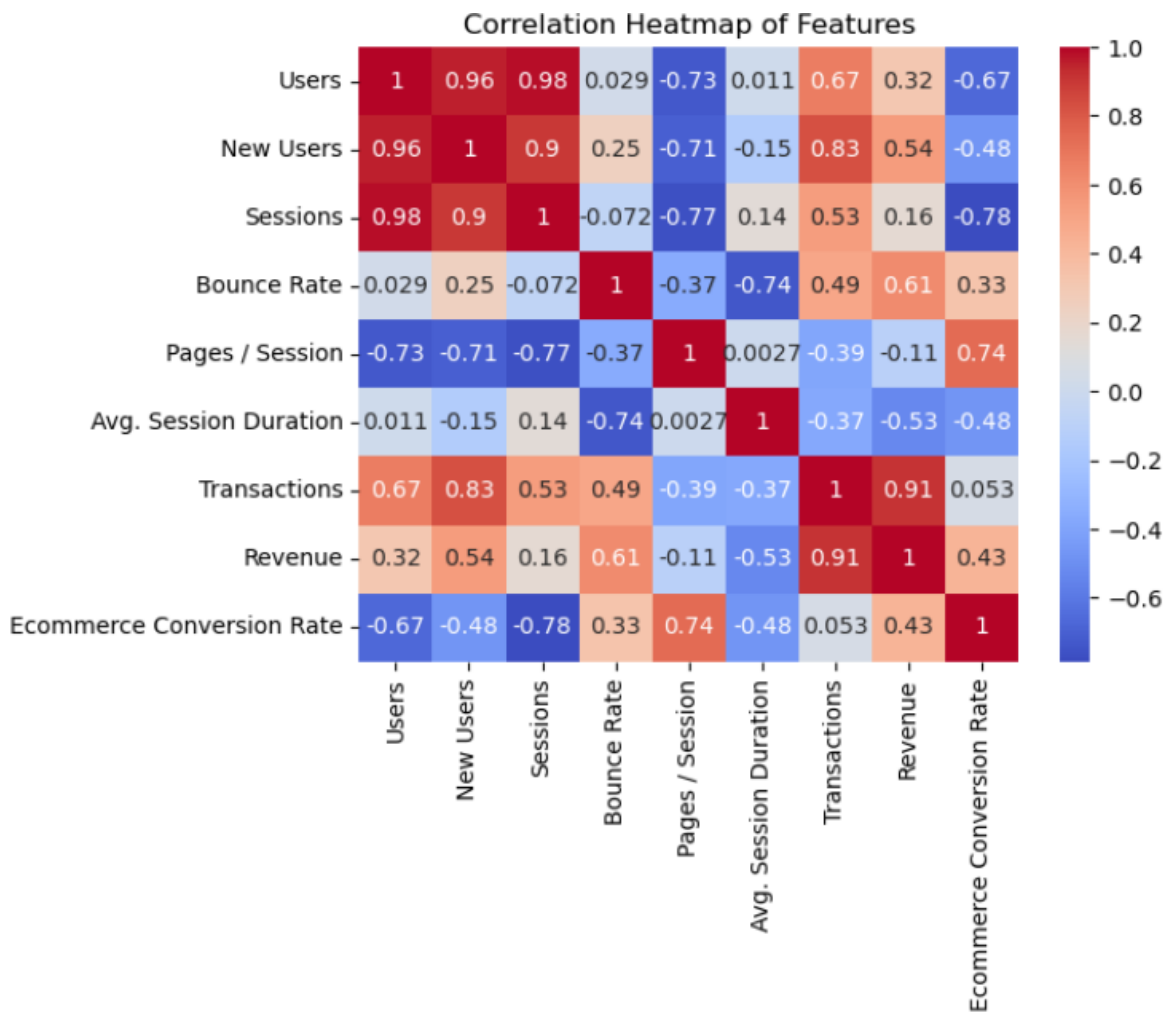


Figure 9: Heatmap showing Correlation between variables

The heatmap we generated shows the relationship between user engagement metrics and e-commerce conversion rates. We found that an increase in the number of users and new users may lead to more sessions and transactions, but a lower conversion rate. Users who spend more time on the website may be less likely to make a purchase, but those who view more pages during a session may be more likely to make a purchase. A higher bounce rate may lead to lower sales. These correlations indicate that user engagement metrics play a critical role in optimizing e-commerce websites for increased sales.

Feature Engineering

We further added engineered features such as Sessions per User, Average Order Value, Revenue per User, and Conversion Rate. These new features were created to provide additional insights into customer behaviour and help inform decision-making processes.

Sessions per User

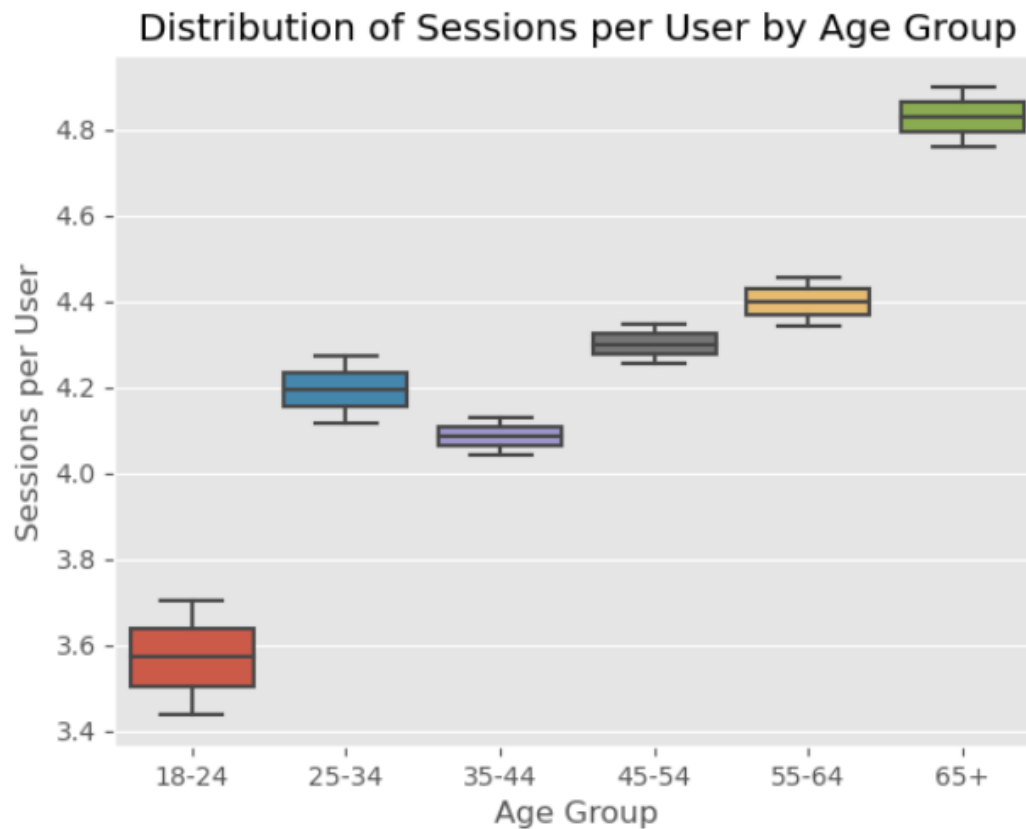


Figure 10: Bar plot of Sessions per User

The sessions per user metric can help business understand how often users are interacting with their website. The dataset reveals that users in the 18-24 age group have the lowest average sessions per user, while users in the 65+ age group have the highest. By analysing this metric, business can create targeted marketing campaigns to increase website engagement for specific age groups.

Average Order Value

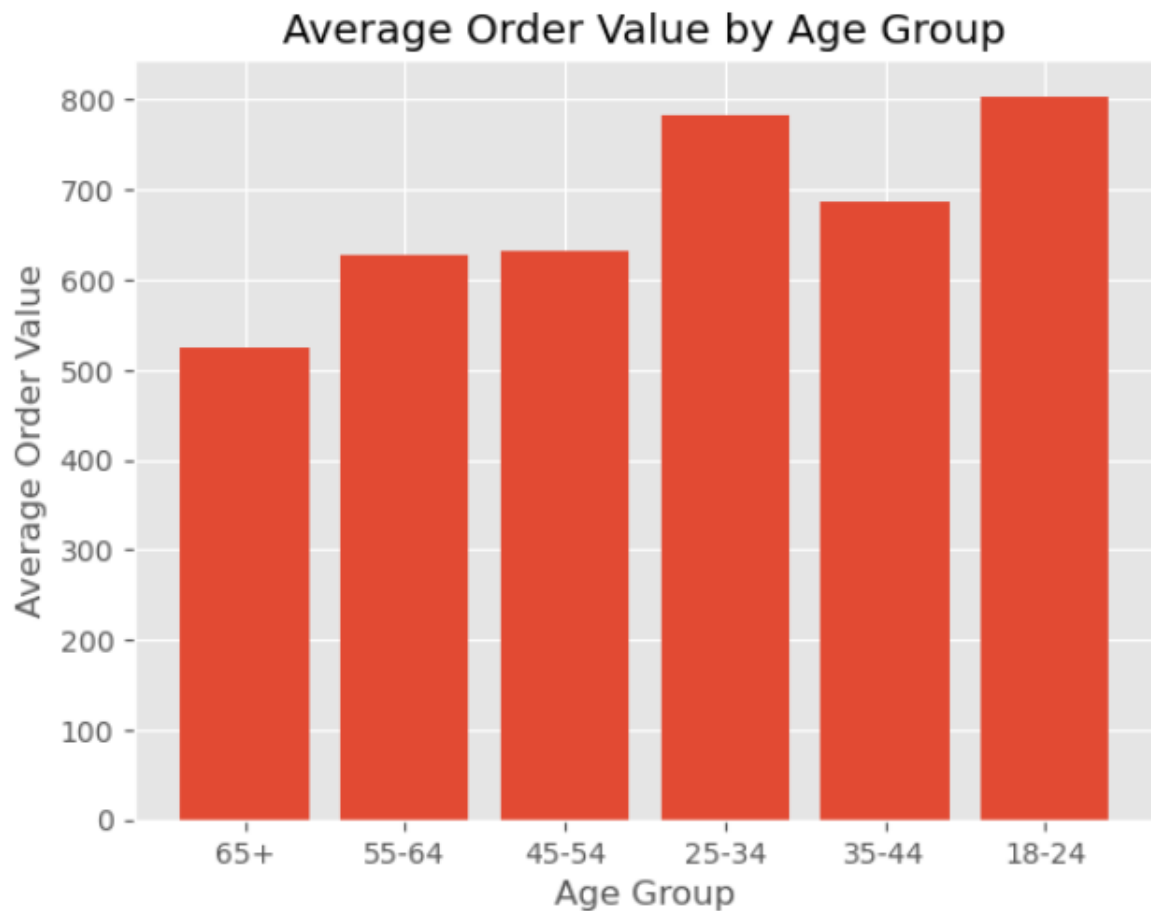


Figure 11: Bar Chart showing Average Order Value

Understanding the average order value can help business evaluate their revenue generation per order. The dataset shows that users in the 18-24 age group have the highest average order value, while users in the 65+ age group have the lowest. Business can create targeted promotions and offers to increase the average order value for specific age groups based on this insight.

Revenue per User

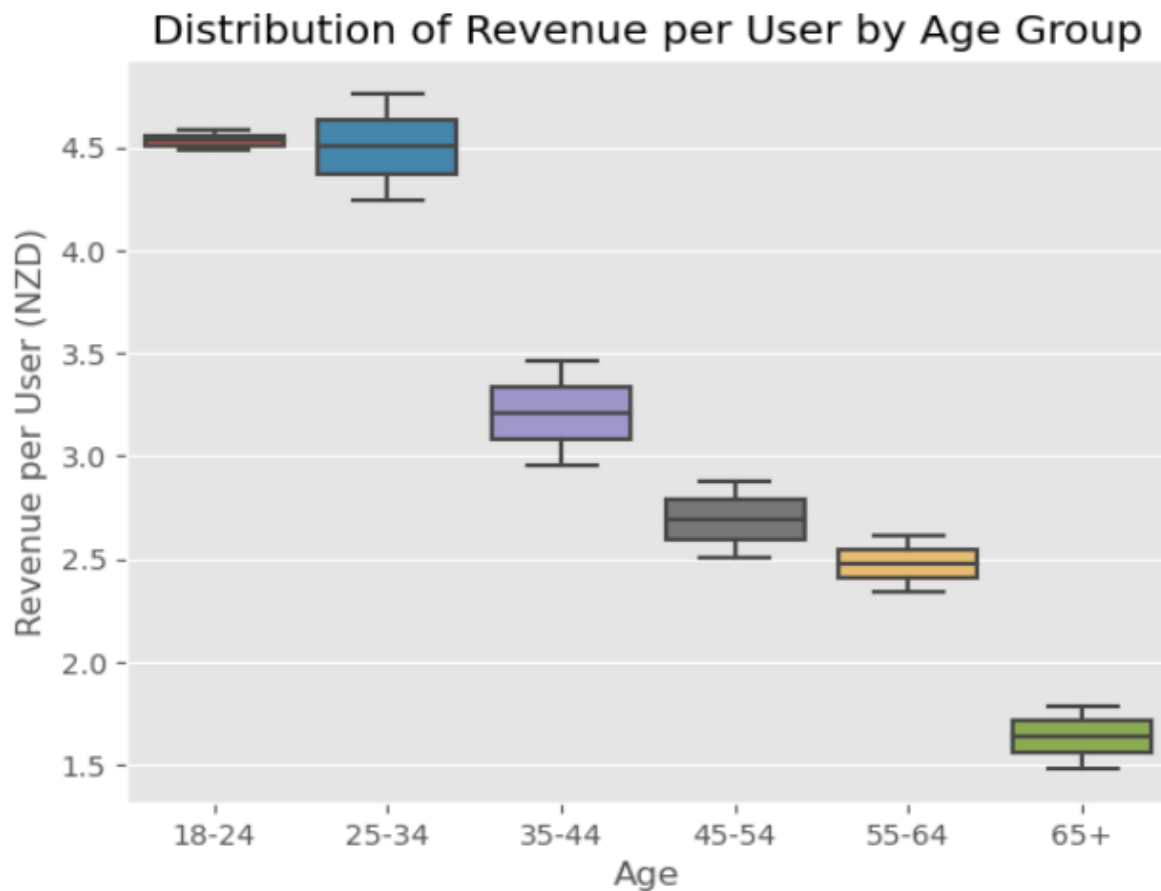


Figure 12: Boxplot showing Revenue per User (NZD)

The revenue per user metric provides a better understanding of how much revenue business generate per user. From the dataset, it is apparent that users in the 25-34 age group generate the highest revenue per user, while users in the 65+ age group generate the lowest. By analysing this metric, business can develop targeted marketing campaigns and promotions to increase revenue per user for specific age groups.

Modelling

We have analysed the performance of three different regression models, namely Random Forest Regressor, Linear Regression, and Support Vector Regression, to gain a better understanding of their effectiveness. Based on our findings,

	Mean Squared Error	R-squared
Random Forest Regressor	0.691	0.524
Linear Regression	0.615	0.155
Support Vector Regression	1.775	-0.461

Random Forest Regressor

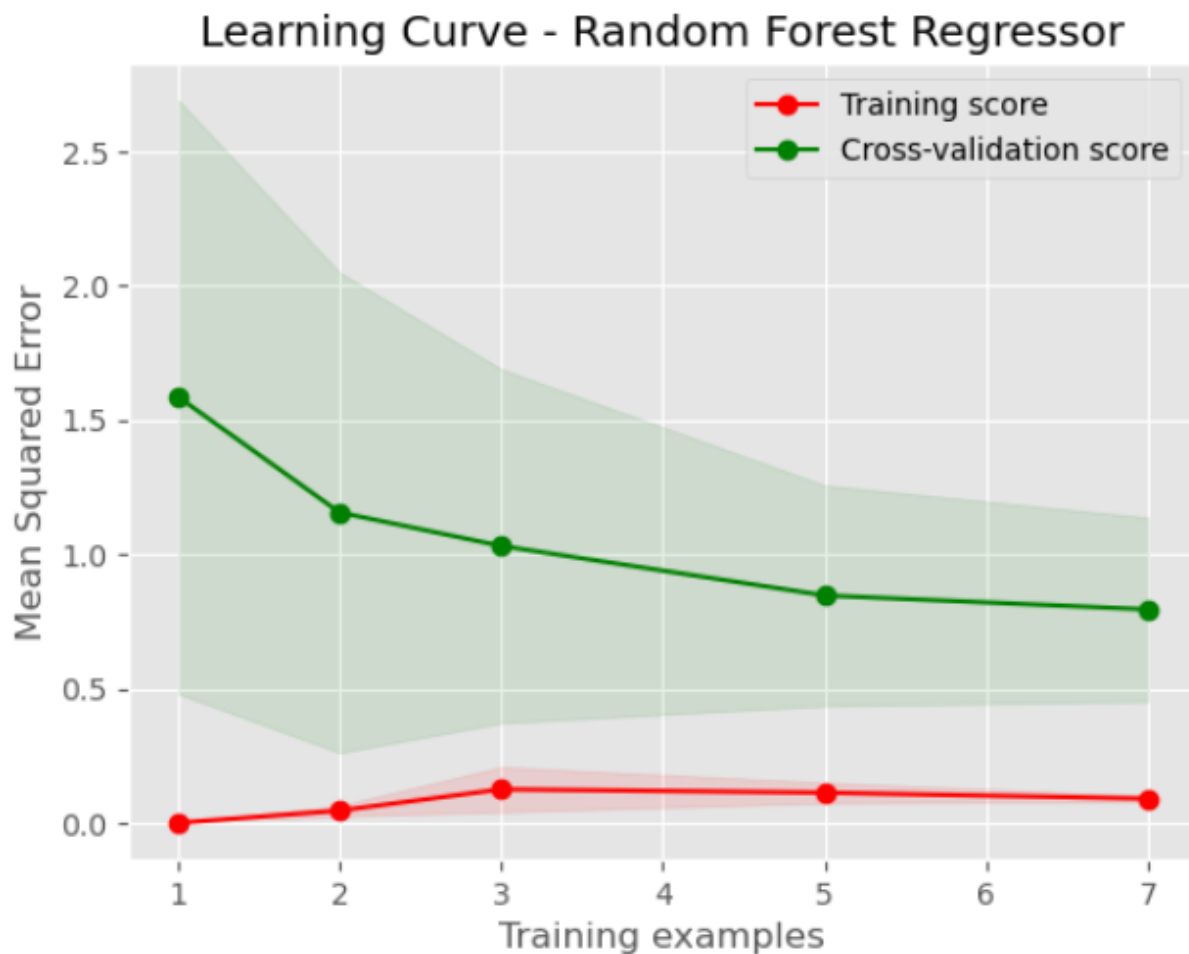


Figure 13: Learning Curve - Random Forest Regression

The Random Forest Regressor has the highest MSE of 0.691, indicating that it has a higher average squared difference between predicted and actual values compared to the other models. This means that the predictions of the Random Forest Regressor model may be less accurate than those of the other models. However, the R-squared value of the Random Forest Regressor is the highest at 0.524, indicating that it can explain a greater proportion of the variance in the dependent variable compared to the other models. This means that the Random Forest Regressor model can better capture the relationship between the independent and dependent variables in the data.

The learning curve for this model indicates that the model is not overfitting or underfitting the data and can generalize well to unseen data.

Linear Regression

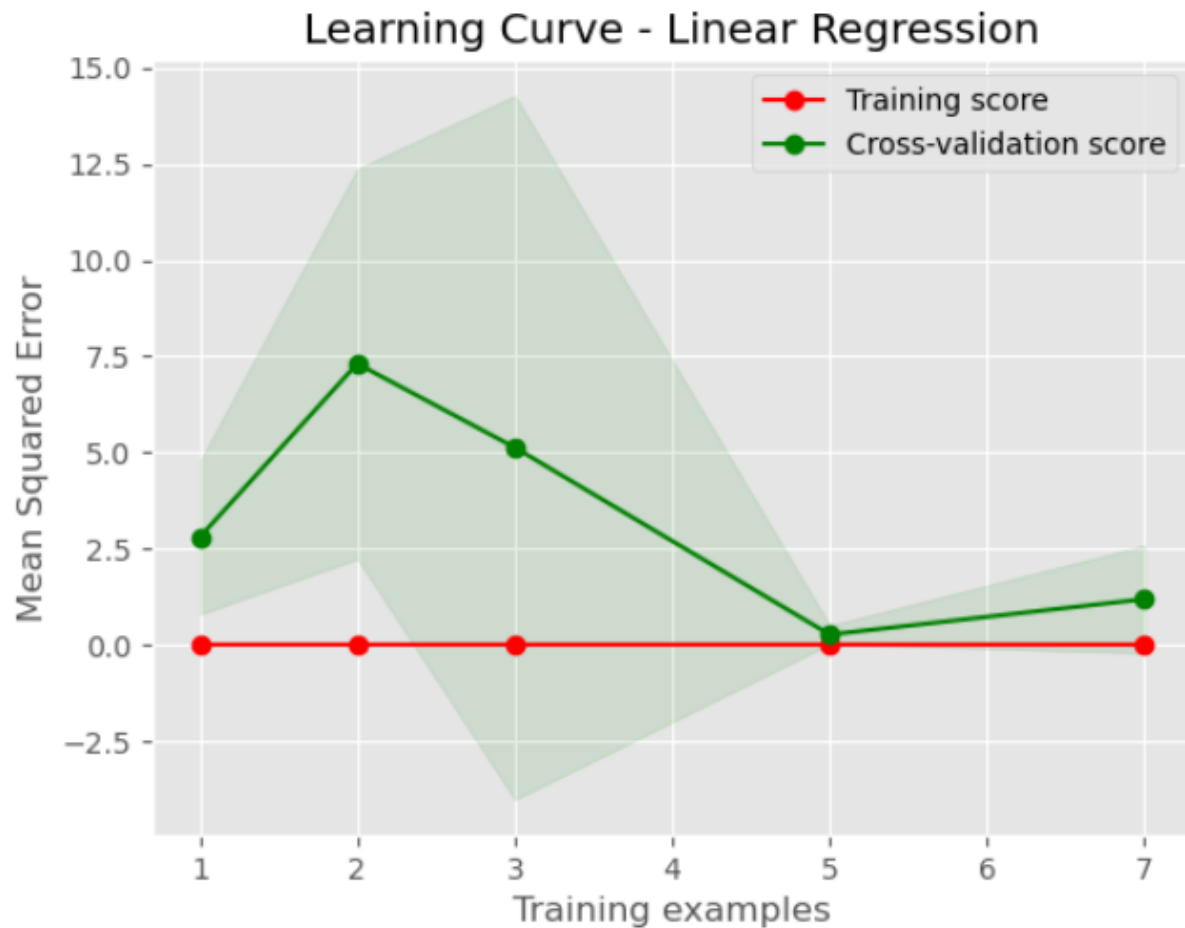


Figure 14: Learning Curve - Linear regression

The Linear Regression model has the lowest MSE of 0.615, indicating that it has the lowest average squared difference between the predicted and actual values among the three models. However, the R-squared value of the Linear Regression model is the lowest at 0.155, indicating that it can only explain a small proportion of the variance in the dependent variable. This means that while the predictions of the Linear Regression model may be more accurate than the other models, it may not capture the complex relationships between the independent and dependent variables as well.

The learning curve suggests that the model is not complex enough to capture the underlying patterns in the data, and more training examples or a more sophisticated model may be needed to improve the performance.

Support Vector Regression

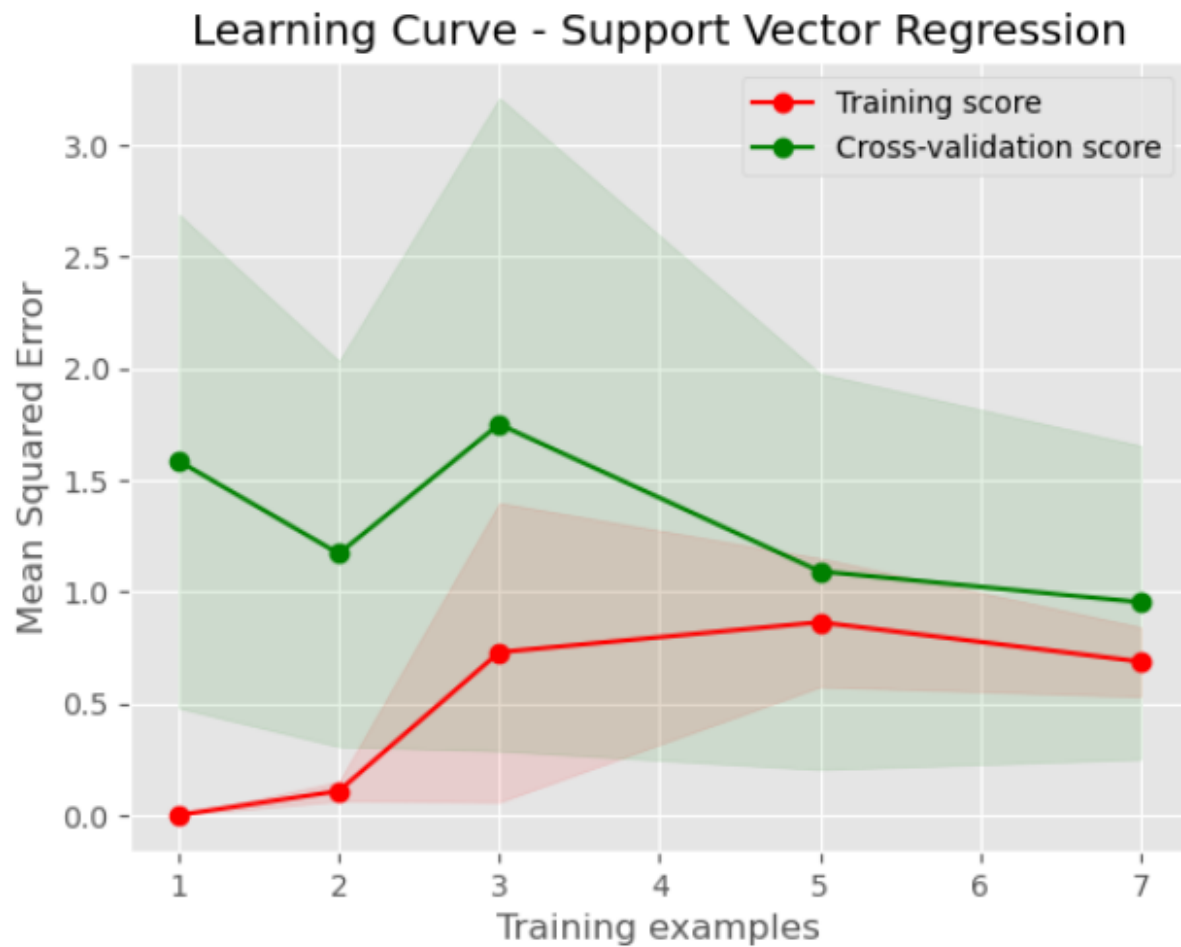


Figure 15: Learning Curve – Support Vector Regression (SVR)

The Support Vector Regression model has the highest MSE of 1.775, indicating that it has the highest average squared difference between the predicted and actual values. Additionally, the R-squared value of the Support Vector Regression model is negative at -0.461, indicating that it is not a good fit for the data and may not be able to explain the variance in the dependent variable.

The learning curve for this model shows that the training and cross-validation scores are declining parallelly as the training examples increase, indicating that the model is not learning from the data and is not able to generalize well to unseen data.

Outcome

Random Forest Regressor model seems to be the best fit for the data based on its high R-squared value, even though it has a higher MSE compared to the Linear Regression model

Implementation

Random Forest Regression can provide valuable insights for business to make informed decisions. It enables business to allocate budgets and devise targeted marketing strategies based on customer groups. By identifying patterns and trends in customer behaviour, business can determine which age and gender groups are most likely to make purchases and which products are more popular among different groups. This information can be used to improve the overall customer experience and develop targeted marketing campaigns. By leveraging the power of Random Forest Regression, business can stay ahead of the competition and make data-driven decisions.

Data answer

By analysing critical variables in the dataset, we have gained insights into how customers engage with our website. These insights have helped us identify areas for improvement and make strategic changes to meet customer needs. Leveraging this data will enhance the user experience and drive brand engagement. We remain committed to optimizing our platform and serving our loyal customers with this valuable information.

Business answer

Our analysis of the provided dataset provides valuable insights that can inform strategies to increase customer purchases. The highest conversion rates were observed among 18-24 year old females and males, suggesting the need to offer products and services that appeal to this demographic and enhance their user experience. Older age groups showed a trend of higher bounce rates, indicating the need for targeted changes to better meet their needs. The platform has a higher conversion rate among females aged 25-44 and males aged 25-34, suggesting the potential for targeted promotions and discounts to drive engagement and improve conversion rates. By tailoring our offerings and strategies to different age and gender groups, we can enhance the user experience, drive greater customer satisfaction and sales.

Response to Stakeholders

To increase our chances of making sales, it's important to understand our customers' engagement patterns. Our analysis of the dataset revealed that the 18-24 and 25-34 age groups, particularly females, have the highest ecommerce conversion rates. To capitalize on this, we suggest creating targeted marketing campaigns and personalized promotions that cater to each demographic. Additionally, we recommend analysing bounce rates and session durations to optimize our website for each group. By taking a data-driven approach, we can better serve our customers and achieve success in our e-commerce efforts.

End-to-end solution

Based on the analysis of the dataset, there are several recommendations that business can follow to optimize their ecommerce strategy and provide a better user experience. By tailoring marketing

efforts to align with each age and gender group's unique needs, business can increase engagement and drive more revenue. Additionally, optimizing website design, content, and user flow can reduce bounce rates and increase average session duration. Finally, by developing a targeted campaign based on demographic analysis, business can provide personalized content and promotions that appeal to their customers.

Overall, by leveraging the insights provided by the dataset, business can improve their ecommerce strategy and provide a more satisfying user experience for their customers.

References

<https://www.smartdatacollective.com/customer-behavior-analysis-telecom-arena/>

<https://www.plecto.com/blog/sales-performance/importance-data-analytics-telecom-industry/>

<https://www.scuba.io/blog/how-to-conduct-a-behavioral-analysis>

<https://www.dataquest.io/blog/learning-curves-machine-learning/>

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html>