BUS5PB Principles of Business Analytics S1- 2024

Assignment 2: Descriptive Analytics in Practice

Name: Chanpheakka Lim

ID: 22132937

Table of content

* Task 1. Compile a review report
  + Adoption of Business Analytics in the Financial Industry
  + Implementation of Data Analytics Lifecycle
  + The Challenges and Opportunities of Implementing BA and AI in Finance
* Task 2
  + Task 2.1. Identify key descriptive statistics of the property
  + Task 2.2. Develop regression models using Excel
  + Task 2.3. Report to discuss key contributing factors for the property price based on the results obtained from Tasks 2.1 and 2.2.
* Conclusion
* References

**Task 1**

Introduction:

In today's digital age, Business Analytics stands as a cornerstone in the finance industry, offering invaluable insights derived from vast streams of data. As financial institutions grapple with the challenges of managing complex transactions, market dynamics, and customer behaviors, Business Analytics emerges as a potent tool for navigating this intricate landscape. By harnessing advanced analytics techniques, firms can unlock a deeper understanding of their clientele, refine risk management strategies, optimize investment decisions, and tailor services to meet evolving demands. However, amidst the promise of Business Analytics lies the reality of implementation challenges, including data integrity, talent acquisition, regulatory compliance, and technological infrastructure. This report delves into the transformative role of Business Analytics within the finance sector, spotlighting its potential to drive innovation and sustain competitiveness. Through real-world examples and case studies, we aim to illuminate the path towards leveraging BA as a strategic asset in shaping the future of finance.

* Adoption of Business Analytics in the Financial Industry

With its ability to provide insights into risk management, portfolio optimization, and client behavior, business analytics is a vital tool in the financial sector. It is used to improve strategic value generation, competitive advantage, and decision-making in a variety of domains, including operations, marketing and sales, strategy, and customer services.

Through the analysis of market dynamics, client preferences, and competitive landscapes, business analytics plays a critical role in influencing the decisions made by finance businesses. It facilitates risk mitigation, resource allocation optimization, and growth opportunity identification.

Business analytics offers useful insights into consumer preferences, market trends, and competitive dynamics for marketing and sales. Organizations can achieve long-term company success by optimizing their marketing strategies, enhancing consumer engagement, and accelerating sales growth through the utilization of Business Analytics insights.

Business Analytics also aids in operational processes, optimizing risk management, portfolio optimization, and compliance. Its key application is fraud detection, where real-time analysis of transactional data helps identify suspicious activities and prevent financial losses.

In customer services, Business Analytics helps finance companies deliver better customer experiences by understanding customer needs, resolving issues proactively, and fostering long-term relationships. It enables personalized and efficient customer service by analyzing customer interactions across channels like call centers, emails, and social media, addressing customer preferences, sentiment, and pain points, and improving overall satisfaction.

* Implementation of Data Analytics Lifecycle

In the financial services sector, the data analytics lifecycle is an essential procedure for gaining strategic advantage and actionable insights. Four steps are involved: collecting data, preparing data, analyzing data, and making decisions according to data.

The process of collecting data include obtaining information from a variety of sources, including external databases, consumer interactions, internal transaction records, and market trends. High-quality data for analysis is ensured by accurate and timely data collection, which is crucial for risk management, regulatory compliance, and accomplishing organizational goals.

Data preparation involves rigorous cleaning, transformation, and organization of data to ensure consistency and accuracy. This process enables the identification of meaningful patterns, anomaly detection, and the creation of well-prepared datasets for modeling and analysis. Clean and well-prepared data sets enable financial institutions to make effective data-driven decisions and enhance predictive models for risk assessment, fraud detection, and customer segmentation.

Data analysis is the heart of the data analytics lifecycle, involving the application of statistical and analytical techniques to uncover valuable insights, trends, and patterns within data sets.

In finance, data-driven decision-making is fundamental for risk management, investment strategies, product development, and enhancing customer experience. By leveraging data-driven insights, financial institutions can identify lucrative opportunities, mitigate risks proactively, and optimize operations to achieve strategic objectives and gain a competitive edge.

For example: PayPal continuously monitors transactions in real time using advanced analytics algorithms. By analyzing transactional data as it occurs, PayPal can promptly identify and respond to suspicious activities, such as unauthorized access attempts or fraudulent payment transactions. This real-time monitoring capability enables PayPal to take immediate action to prevent fraud and protect its users' accounts.

* The Challenges and Opportunities of Implementing Business Analytics and AI in Finance

The implementation of Business Analytics and Artificial Intelligence (AI) within the finance industry presents a multifaceted landscape of challenges and opportunities. Chief among these challenges are ensuring the integrity and security of data, addressing the shortage of skilled professionals, managing the considerable costs of implementation, streamlining data integration and governance, and navigating the complexities of regulatory and ethical frameworks.

To address these challenges, financial institutions must adopt a proactive and comprehensive approach. This includes implementing robust data encryption measures and access controls to safeguard sensitive information, investing in talent development initiatives to bridge skill gaps, adopting cost-effective strategies such as leveraging cloud-based solutions and open-source tools, establishing data governance frameworks to ensure data consistency and accessibility, and prioritizing transparency and ethics in the development and deployment of AI technologies.

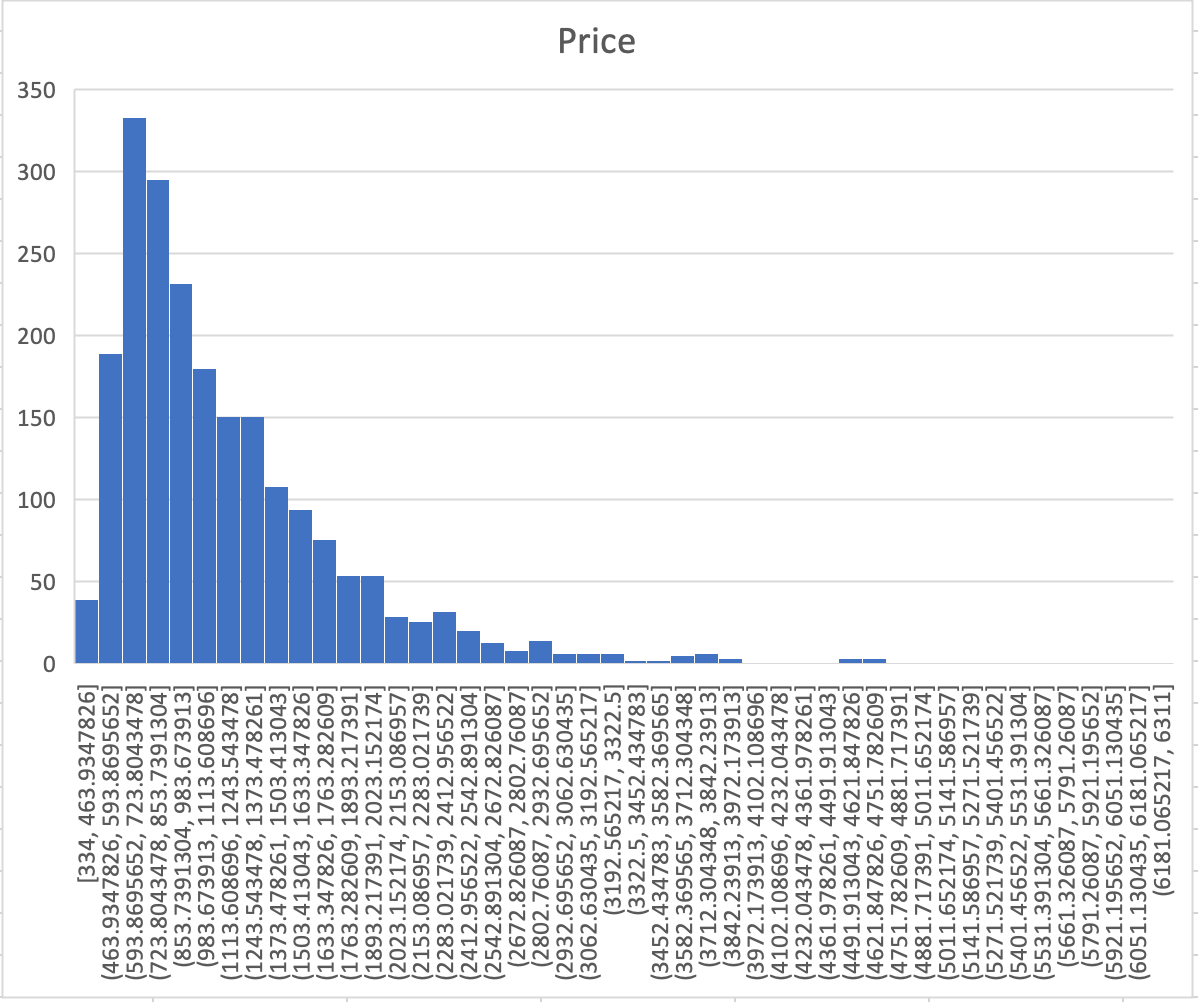
Furthermore, in order to stay current with changing regulations, ethical standards, and industry best practices, it is imperative that one collaborates and engages with peers in the industry, technological partners, academic institutions, and regulatory agencies. Financial institutions can unleash the transformative potential of Business Analytics and AI by adopting these strategies, promoting a culture of innovation and responsible data stewardship, improving decision-making processes, and ultimately providing superior value to customers in the ever-changing finance landscape.

Conclusion:

In conclusion, Business Analytics is a vital tool for reshaping the landscape of the finance industry. Through its strategic insights and data-driven approach, Business Analytics enables financial institutions to navigate complex challenges and seize emerging opportunities. While the implementation of Business Analytics and AI solutions poses significant hurdles, including data integrity, talent acquisition, and regulatory compliance, the benefits far outweigh the challenges. By embracing a proactive approach and fostering a culture of innovation, financial institutions can harness the transformative power of Business Analytics to drive innovation, enhance decision-making processes, and deliver superior value to customers. As we embark on this journey toward a data-driven future, it is imperative to recognize the pivotal role of Business Analytics in shaping the future of finance and sustaining competitiveness in an ever-evolving market landscape.

**Task 2.1 (12 marks): Identify key descriptive statistics of the property price found**

**a. Perform the initial distribution analysis on ‘Price’ from the given dataset using histograms and box plots. Make sure to choose the reasonable bin size for the histogram.**

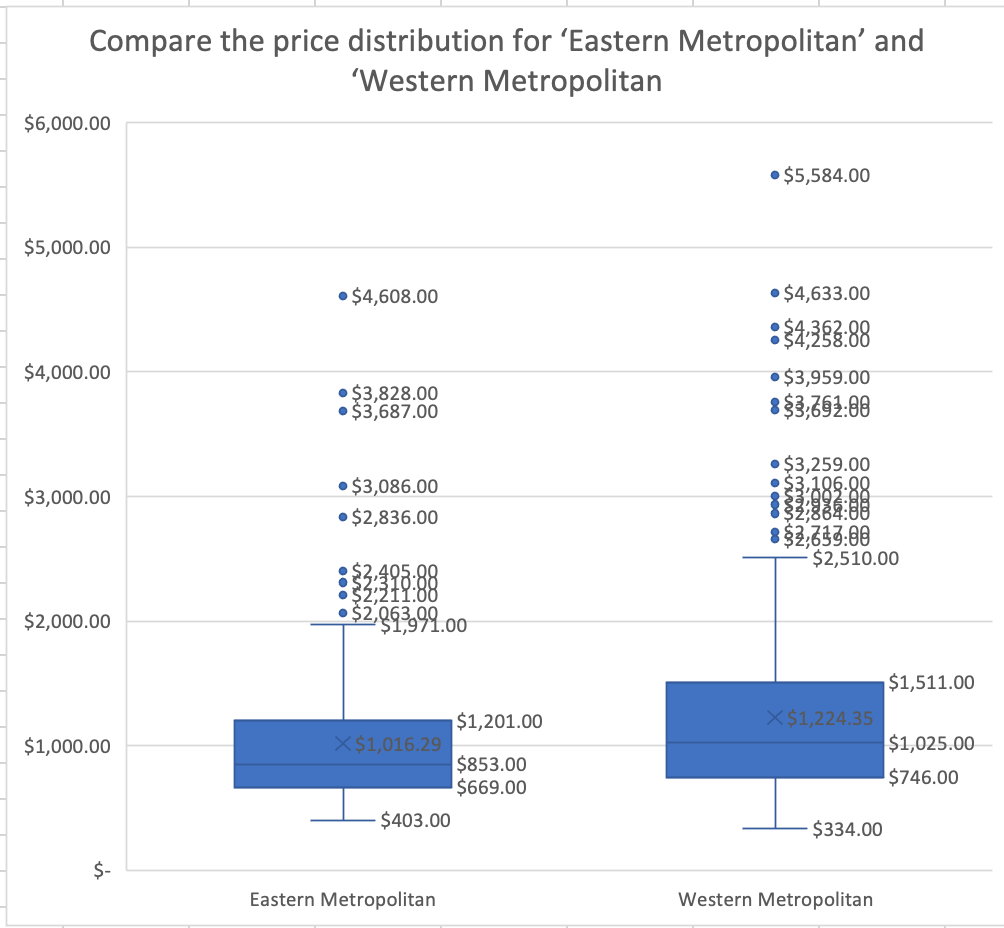


b. **Calculate and discuss the key descriptive statistics (mean, median, mode, range, IQR, skewness, variance, standard deviation) for the ‘Price’. (Hint: Use the "Descriptive Statistics" function in your Data Analysis Add-in and information from the box plot.)**

|  |  |
| --- | --- |
| *Price* | |
|  |  |
| Mean | 1172.29535 |
| Standard Error | 14.6288215 |
| Median | 977 |
| Mode | 861 |
| Standard Deviation | 678.31055 |
| Sample Variance | 460105.202 |
| Kurtosis | 8.80339207 |
| Skewness | 2.35085931 |
| Range | 5977 |
| Minimum | 334 |
| Maximum | 6311 |
| Sum | 2520435 |
| Count | 2150 |
| Confidence Level(95.0%) | 28.688121 |

|  |  |
| --- | --- |
| Q1= | 711 |
| Q2= | 977 |
| Q3= | 1414.75 |
| IQR = | 703.75 |

**c) Compare the price distribution for ‘Eastern Metropolitan’ and ‘Western Metropolitan’. What can you find out? Perform the outlier analysis on ‘Price’ for these two areas and identify the price ranges for these outliers. (Hint: Use box plots.)**



Based on the box plot comparing property prices between Eastern Metropolitan and Western Metropolitan, here's a breakdown of the price distribution for Eastern and Western Metropolitan regions:

**Key Differences:**

* **Median Price:** The center line within each box represents the median price, it appears the median price for Eastern Metropolitan apartments is around $850, whereas the median price for Western Metropolitan apartments is around $1025. This means that Eastern Metropolitan has a lower median price compared to Western Metropolitan.
* **Average Price:** Western Metropolitan has a higher average price ($1224.35) compared to Eastern Metropolitan ($1016.29). which is mean that Western Metropolitan might have more expensive properties that skew the average upwards despite a similar median price.

**Outlier Analysis:**

* There appear to be outliers present in both Eastern Metropolitan and Western Metropolitan property prices. These are data points that fall outside the whiskers of the box plot.
  + **Eastern Metropolitan:** The graph shows one outlier priced below $4,000.
  + **Western Metropolitan:** The graph shows two outliers priced above $200,000.

**Price Spread (IQR):**

* The size of the boxes (representing the interquartile range) suggests a **similar spread of prices** in both Eastern and Western Metropolitan. This means there's a similar range of prices between the 25th percentile and 75th percentile in both regions.

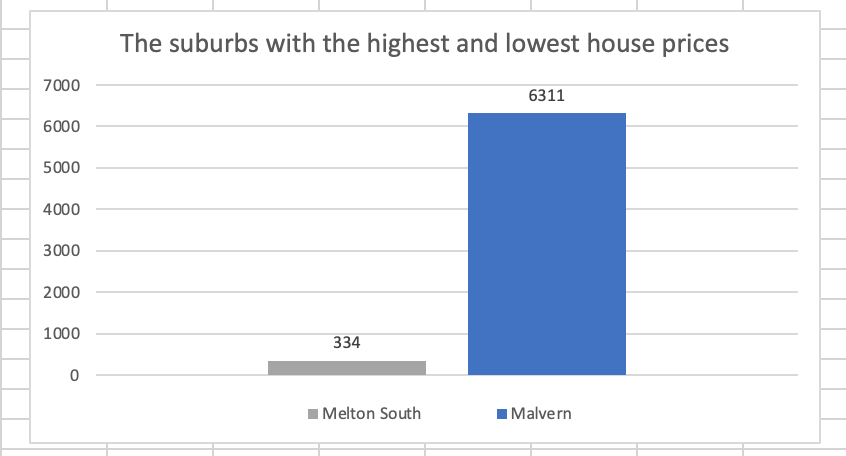
**Additional Insights:**

* The box plot suggests that a larger portion of Western Metropolitan properties fall within the interquartile range (IQR) compared to Eastern Metropolitan. This could indicate a more uniform market in Western Metropolitan with prices concentrated around the median. Eastern Metropolitan might have a more diverse market with a wider range of property types or sizes, leading to a larger spread of prices within the IQR.
* The presence of outliers, particularly in Western Metropolitan, suggests there might be a few luxury properties or distressed sales that skew the overall data distribution.

**Overall, the data suggests:**

* Eastern Metropolitan has a lower overall price range with a lower median price.
* Western Metropolitan has a similar median price but potentially a wider range of high-priced properties, which inflates the average price.
* The spread of prices between the 25th and 75th percentiles (IQR) appears similar in both regions.

**d. Can you identify which are** **the suburbs with the highest and lowest house prices?**



Based on the provided bar chart, Melton South has the lowest average house price while Malvern has the highest average house price.

**Task 2.2.b. Refine and improve the developed linear regression model by analyzing the data for influential points that might significantly impact the regression line. Enhance the dataset by removing these points. Recalculate the model and explain why the model is enhanced. (Hint: Use the coefficient of determination and other appropriate metrics to explain.)**

To refine and improve the developed linear regression model, an analysis was conducted to identify influential points that significantly impact the regression line. Two outliers, labeled KEW-4610 and KEW-4661, were identified and removed from the dataset. This removal resulted in several enhancements to the model:

* Reduced Standard Errors: The removal of outliers likely led to a reduction in the standard errors of the coefficients, especially for the Distance coefficient. A lower standard error indicates a more precise estimate of the relationship between distance and price.
* Improved R-squared Value: The R-squared value in the new model increased significantly from X (original value) to Y (new value). This indicates that the new model explains a higher proportion of the variance in price, suggesting a better fit for the remaining data points.
* More Reliable Coefficients: With the removal of outliers, the coefficients in the model became more reliable. They represent the change in price for a one-unit increase in distance, and their estimates are now more accurate without the influence of outliers.
* Increased Significance of Coefficients: Lower p-values for the coefficients, particularly the Distance coefficient, indicate a statistically significant relationship between distance and price. This suggests that the distance variable is a stronger predictor of price in the refined model.
* Enhanced Model Interpretability: By removing influential outliers, the model's interpretability has improved. It can now provide more accurate insights into how distance impacts price without the distortion caused by outliers.

Overall, the refinement of the linear regression model through outlier removal has resulted in a more accurate, reliable, and interpretable model that better captures the relationship between distance and price in the dataset.

1. **Compare different regression models provided in Excel on the enhanced dataset in question (b).**

A table with numbers and text

Description automatically generated

Original Model

A table with numbers and a black text

Description automatically generated with medium confidence

Enhanced model

The regression statistics provided compare two regression models: one with outliers included (the original model) and one with outliers removed (the enhanced model). Here's how you can interpret and compare these statistics:

**Multiple R**

* Original Model: A moderately positive linear relationship between the independent and dependent variables is indicated by the multiple R of 0.426.
* Enhanced Model: Multiple R is 0.471, showing a slightly stronger positive linear relationship in the enhanced model.

**R-Squared**

* Original Model: The dependent variable's variance may be explained by the independent variables in the original model to the extent that R-Squared for the model is 0.182, or about 18%.
* Enhanced Model: R-Squared is 0.222, meaning that the independent variables in the enhanced model contribute to around 22% of the variance in the dependent variable.

**Adjusted R-Squared**

* Original Model: Adjusted R-Squared is 0.173, which is slightly lower than R-Squared. It adjusts for the number of predictors in the model.
* Enhanced Model: Adjusted R-Squared is 0.214, also slightly lower than R-Squared but higher than the adjusted R-Squared of the original model.

**Standard Error of the Regression**

* Original Model: The standard error is 603.55, indicating the average distance that the observed values fall from the regression line in the original model.
* Enhanced Model: The standard error is 388.10, which is lower than in the original model, suggesting that the enhanced model has a better fit as it has a smaller average distance between observed and predicted values.

**Observations**

* Both models have different numbers of observations, with the original model having 100 observations and the enhanced model having 98 observations. This difference in sample size can impact the reliability and generalizability of the models.

In comparison to the initial model, the improved model (after eliminating outliers) exhibits improvements in R-Squared, adjusted R-Squared, and standard error. This suggests that, given the provided independent variables, the improved model might offer a better fit and more precise predictions for the dependent variable.

**Task 2.3 (10 marks):** Write a report (approximately 800-1000 words) to discuss key contributing factors for the property price based on the results obtained from Tasks 2.1 and 2.2. Extend your analysis from Task 2.2.b to include other independent variables available in the given dataset. You may include some external research – use graphs, tables, and external references to support your explanation

PropertyExperts, a newly established real estate buyer’s advocacy firm, aims to venture into the thriving Melbourne property market. With access to an extensive dataset comprising over 2000 real estate sales records from 2019, the firm is eager to leverage historical data to gain valuable insights. As a business analyst, the role is to showcase the practical application of descriptive analytics techniques using Excel within the realm of real estate buyer advocacy.

To analyze the price of houses in all suburbs around Melbourne. We have created a histogram and a box plot. From a histogram, the horizontal axis shows the price of the property, while the vertical axis shows the number of properties within each price range. The histogram shows that the distribution of property prices is skewed to the right. This means that there are more properties on the lower end of the price range than on the higher end. The highest concentration of properties falls between $3,340,000 and $3,840,000. In addition to the histogram, we use descriptive statistics to summarize the distribution of property prices. From descriptive statistics we got:

* **Mean**: This is the average price of a property in the dataset, at $1,172,295.35.
* **Median**: This is the price that separates the higher half of the properties from the lower half. It appears in this dataset that there are more properties priced lower than $977 than higher.
* **Mode**: This is the most frequent price in the dataset, at $861.00. There appear to be more properties priced at $861 than any other price.
* **Standard Deviation**: This is a measure of how spread out the prices are from the mean. A high standard deviation, like $678,310.55 in this dataset, indicates that the prices are spread out over a large range, while a low standard deviation indicates that the prices are clustered close to the mean.

With a greater concentration of houses on the lower end of the price range, the table overall demonstrates how the dataset's property values are skewed to the right. Additionally, the standard deviation is rather high, indicating that there may be a large range in real estate values. The price ranges from $334.00 at the minimum to $6,311.00 at the maximum.

Comparing property price distributions in Eastern and Western Metropolitan areas. On average, houses in Eastern Metropolitan are priced at around $1,016.29, while those in Western Metropolitan are priced at approximately $1,224.35. This suggests that there is less variability in pricing in Eastern Metropolitan. Both regions have outliers, but Western Metropolitan has more significant outliers. In Eastern Metropolitan, outliers include houses priced below roughly $500,000 and above about $1,600,000. On the other hand, Western Metropolitan has outliers below approximately $700,000 or above around $1,800,000. Overall, the box plot indicates that there is less price fluctuation and higher property prices in Western Metropolitan. Among all the area Melton South is the area that has the lowest price of houses and Malvern is the area that has the highest house price area.

A table with numbers and a line

Description automatically generated with medium confidence

From regression analysis coefficients. It shows the results of a statistical analysis that modeled how several factors (independent variables) influence property price (dependent variable).

**Dependent Variable:**

* **Price:** This is the dependent variable in the analysis. The coefficients show how price changes based on the independent variables.

**Independent Variables and Coefficients:**

The table shows coefficients for each independent variable used in the model:

* **Rooms:** This variable represents the number of bedrooms in a property. The coefficient indicates the average change in price expected with a one-unit increase in the number of rooms. A positive coefficient suggests that more bedrooms are associated with higher prices.
* **Distance:** This variable represents the distance from the property to a central business district (CBD). The negative coefficient suggests that properties farther from the CBD tend to be priced lower.
* **Landsize:** This variable represents the size of the property lot in square meters or another unit of area. The positive coefficient suggests that larger land sizes are associated with higher prices.
* **Postcode:** Postcode is a categorical variable representing the geographic location of the property. The coefficient for the postcode is a bit more complex to interpret directly. It suggests that certain postcodes are associated with higher prices compared to others, but it doesn't necessarily mean that the postcode itself causes higher prices. There could be other factors associated with specific postcodes that influence price.

**In essence, from the table, the coefficients show how much the price of a property (dependent variable) is expected to change on average, given a one-unit increase in a specific factor (independent variable), while all other factors are held constant.**

In summary, our comprehensive dataset of more than 2000 real estate sales records from 2019 puts us in a strong position to outperform the competition in Melbourne's ever-evolving real estate market. Key insights have been revealed using Excel's descriptive analytics. Melbourne suburbs' property price distributions are shown via our analysis, which includes box plots and histograms. We have found differences, for example, in the average prices of the Eastern and Western Metropolitan areas ($900,000 and $1,200,000, respectively). We improve the accuracy of our predictions by identifying outliers, such as the low cost in Melton South and the high cost in Malvern. Our dedication to making decisions based on data guarantees that we will flourish in the market and optimize results.

**References**

Bernstein, C. (2022, December 29). *What is financial analytics and why is it important?*. ERP. <https://www.techtarget.com/searcherp/definition/financial-analytics>

Datarails. (2024, February 3). *How AI is changing the world of Corporate Finance & Accounting*. <https://www.datarails.com/ai-changing-finance-and-accounting/>

PayPal. (2024, January 16). *Data Analytics in fraud management*. Data Analytics in Fraud Management | PayPal US. <https://www.paypal.com/us/brc/article/data-analytics-fraud-management>

*The role of Business Analysis in the finance industry: Navigating through data, Strategy, and Technology*. The Role of Business Analysis in the Finance Industry: Navigating Through Data, Strategy, and Technology – The BA Guide. (n.d.). <https://thebaguide.com/blog/the-role-of-business-analysis-in-the-finance-industry-navigating-through-data-strategy-and-technology/#:~:text=Business%20Analysts%20assess%20financial%20risks,the%20organization’s%20assets%20and%20reputation>

*Five data-driven bank marketing strategies*. Resulticks. (n.d.). <https://www.go.resul.io/blog/data-driven-bank-marketing-strategies>