**BANASD602 Visual Analytics and Storytelling**

Name

Student ID

Email Address

Learning Facilitator

Submission date

**Introduction**

In today's world, data has become crucial for businesses in deciding on their next move. Data collected by applications and smart devices is being utilized to drive visions and steps for institutions that want to impact and have a competitive edge (Banachewicz, & Massaron 2022). This report examines one such business data by transforming the raw data into something meaningful through visual analytics. By leveraging visualization software such as Tableau, the analysis proposes to evaluate where the business is thriving and its shortcomings.  The storytelling explored in this paper is guided by investigative questions to uncover trends and patterns that ultimately drive decisions.

**Overview**

The dataset being analyzed is sourced from Kaggle <https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset?resource=download>. It contains attributes such as transactional records, customer demographics, product categories, payment methods, and shopping mall locations between 2021 and 2023. From this data set, the task is to understand consumer behavior, sales trends, and purchasing patterns across different demographics and regions. Upon downloading the data set there were inconsistencies and errors, such as values exceeding 100 years. These were filtered out and others were flagged to ensure accuracy.  Attributes used in the visualizations to determine trends include demographics, such as age and gender. These two attributes would inform the business's marketing team who to target and based on which products.  The product category and sales attribute are essential in understanding the items generating high and low sales and the trends in buying in the three years.  Tableau, a visualization software, was employed to show the trends in the data and what it meant for the business.

**Data Preparation**

  Before analysis, the raw data must be prepared to achieve accuracy in interpreting the results.  The cleaning process entailed removing duplicates and handling the null values (Shan & Gubin, 2019).  It was also filtered to check whether there were inconsistencies per attribute. The null values were replaced with appropriate defaults and standardizing variables. For instance, "tech" and "electronics" were tied to one category to help improve the data quality. The next check was on the data types on each attribute to be set correctly, such as ensuring price and sales are set to currency and age is set to numbers. The next step was grouping the demographics like age into age groups that would give better insights (Afzal et.al., 2023). The preparation also entailed calculating the sales using the unit price and quantity to know which mall, category, or season was performing well and which was struggling.  Lastly, the data was prepared by extracting month and year from the date provided to analyze sales trends by season throughout the year.

**Findings**

The target audience for this analysis varies based on the interpretation. Business owners and shopping mall managers would benefit from this analysis to understand sales trends by category or demographics. This information helps optimize store placement and improve sales. Marketing teams would also find the analysis invaluable based on customer trends and seasonal fluctuations. Their interest would be to know where to direct their targeted promotions and marketing strategies.  Lastly is the inventory and supply chain team whose primary interest is to manage the stocking of products to avoid overstocking or vice versa.

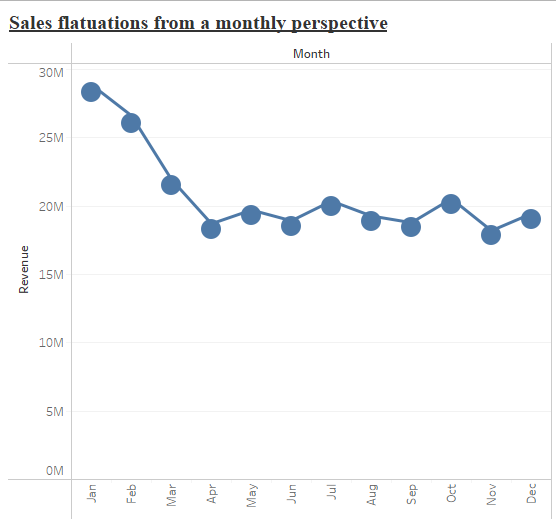
In the investigative question of which product generates the highest revenue, clothing ranked top with about $113m in sales. The shoes and technology products were the second and third highest revenue-generating at $66m and $ 57m, respectively. These categories likely attract frequent purchases due to fashion trends and technological advancements. In contrast, Food & Beverage, Books, and Souvenirs recorded the lowest sales suggesting lower consumer interest or price points. Increasing marketing efforts for underperforming categories or bundling them with high-selling products could enhance overall sales. A graph with blue bars

AI-generated content may be incorrect.

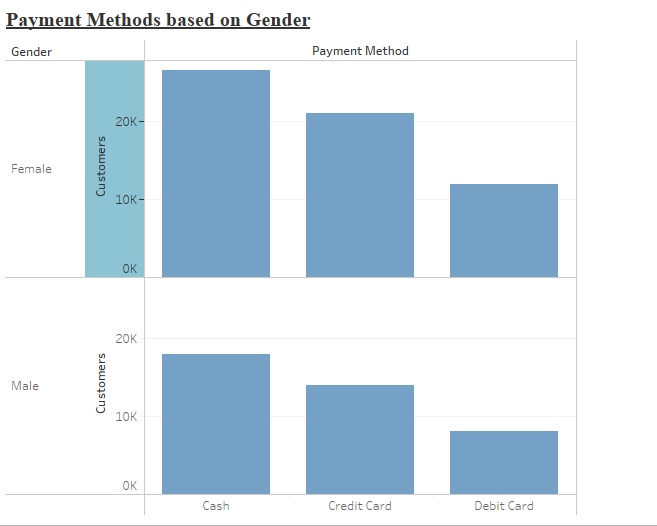
 An analysis of sales trends on different demographics showed females accounted for higher sales, generating approximately $150 million, compared to $101 million from males. This suggests that women may shop more frequently or spend more per transaction. On the other hand the analysis based on age showed that individuals in the age group 30- 44 generated the highest revenue at $73 million, followed closely by the 45-59 age group with 71 million. As depicted in the chart below middle-aged consumers are key sales drivers, most likely due to disposable income. The 15-29 age group ranked third with sales of $58 million, possibly reflecting budget-A screen shot of a graph

AI-generated content may be incorrect.conscious younger shoppers, while the 60-74 group had the lowest sales at 48 million. A screenshot of a graph

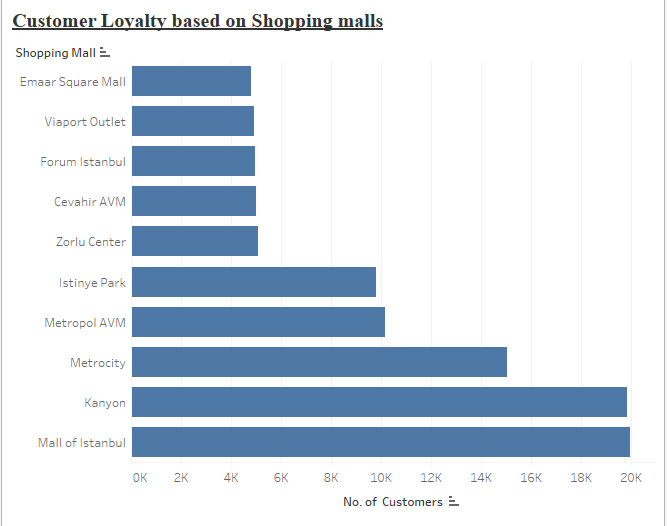
AI-generated content may be incorrect.

The analysis of seasonal sales fluctuations was based on the months of the three years. The visualization chart reveals distinct trends in customer footfall. January, February, and March recorded the highest sales, with a steady decrease highlighting a strong start to the year. However, April saw a decline in sales ($18M) from ($21m) million in March, followed by a slight recovery in May (19.7M) before dropping again in June. Sales peaked in July with ($20.3M) in sales but dipped again in August ($19.2M) and September ($18.7M). A notable rise occurred in October ($20.5M), followed by a decline in November ($18.2M) before a slight rebound in December ($19.5M). 

An analysis of prevalent payment methods across different demographics indicates that cash is the most preferred payment method among males and females. This trend suggests that cash transactions remain dominant despite the availability of digital payment options. The pattern is also seen when evaluating age-wise. Cash usage ranked highest across all age groups, with credit cards ranking second and debit cards third. This preference for cash could be influenced by convenience and long-standing financial habits.



Evaluating customer retention across shopping malls reveals that the Mall of Istanbul and Kanyon ranked top with 19,943 and 19,823 customers respectively. This means these two malls have gained a competitive edge in the market that attracts and retains shoppers. Further studies should establish what drives loyalty in these malls to help other shopping centers adopt similar approaches to enhance customer retention.



**Conclusion**

In conclusion the analysis provides valuable insights into customer behavior, sales trends, and payment preferences. The visualization tool was essential in identifying peak sales periods, high-performing product categories, and customer loyalty patterns can help businesses make data-driven decisions. By leveraging these insights the target audience such as business owners and marketing teams is able to enhance overall profitability and growth.

**References**

Afzal, S., Ghani, S., Hittawe, M. M., Rashid, S. F., Knio, O. M., Hadwiger, M., & Hoteit, I. (2023). Visualization and visual analytics approaches for image and video datasets: A survey. *ACM Transactions on Interactive Intelligent Systems*, *13*(1), 1-41. https://doi.org/10.1145/3576935

Customer shopping dataset - Retail sales data. (n.d.). Kaggle: Your Machine Learning and Data Science Community. Retrieved March 3, 2025, from <https://www.kaggle.com/datasets/mehmettahiraslan/customer-shopping-dataset?resource=download>

Shan, H., & Gubin, E. I. (2019). Data cleaning for data analysis. In *Молодежь и современные информационные технологии: сборник трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых учёных, 3-7 декабря 2018 г., г. Томск.—Томск, 2019.* (pp. 387-388).

Banachewicz, K., & Massaron, L. (2022). *The Kaggle Book: Data analysis and machine learning for competitive data science*. Packt Publishing Ltd.