om bangale

48005266

Assignment 2

Using Snowflake for making Visualisations and Predictive Modelling on BBE Dataset

# Executive Summary

This report explores how Snowflake was used for both predictive modelling and dashboard creation, showcasing its strength as a flexible and efficient data warehousing platform. By building predictive models directly within Snowflake, we forecasted sales trends based on product data. Dashboards were also developed within Snowflake, with options for seamless integration into external tools like Tableau and Jupyter notebook for enhanced visualizations and advanced model deployment. Additionally, we compare Snowflake to other popular data warehousing solutions—AWS Redshift, Google BigQuery, and Microsoft Azure Synapse—to highlight Snowflake's advantages in scalability, user-friendliness, and integration capabilities.

# Introduction

Best Bike Ever (BBE) tasked our team with applying predictive analytics and visualizations using a scalable data platform to improve their decision-making process. Snowflake was chosen for this project due to its robust cloud-native infrastructure, which excels in efficient data storage, processing, and analytics.

The objective was to build a predictive model to forecast sales based on important product features, and then develop a dashboard to present these insights visually. Both the predictive model and dashboard were created directly in Snowflake, demonstrating its ability to handle complex data analytics efficiently. Additionally, we explored how Snowflake integrates with external tools like Tableau for real-time visualizations and Jupyter for more sophisticated machine learning tasks through live connections.

Username: bangale0824

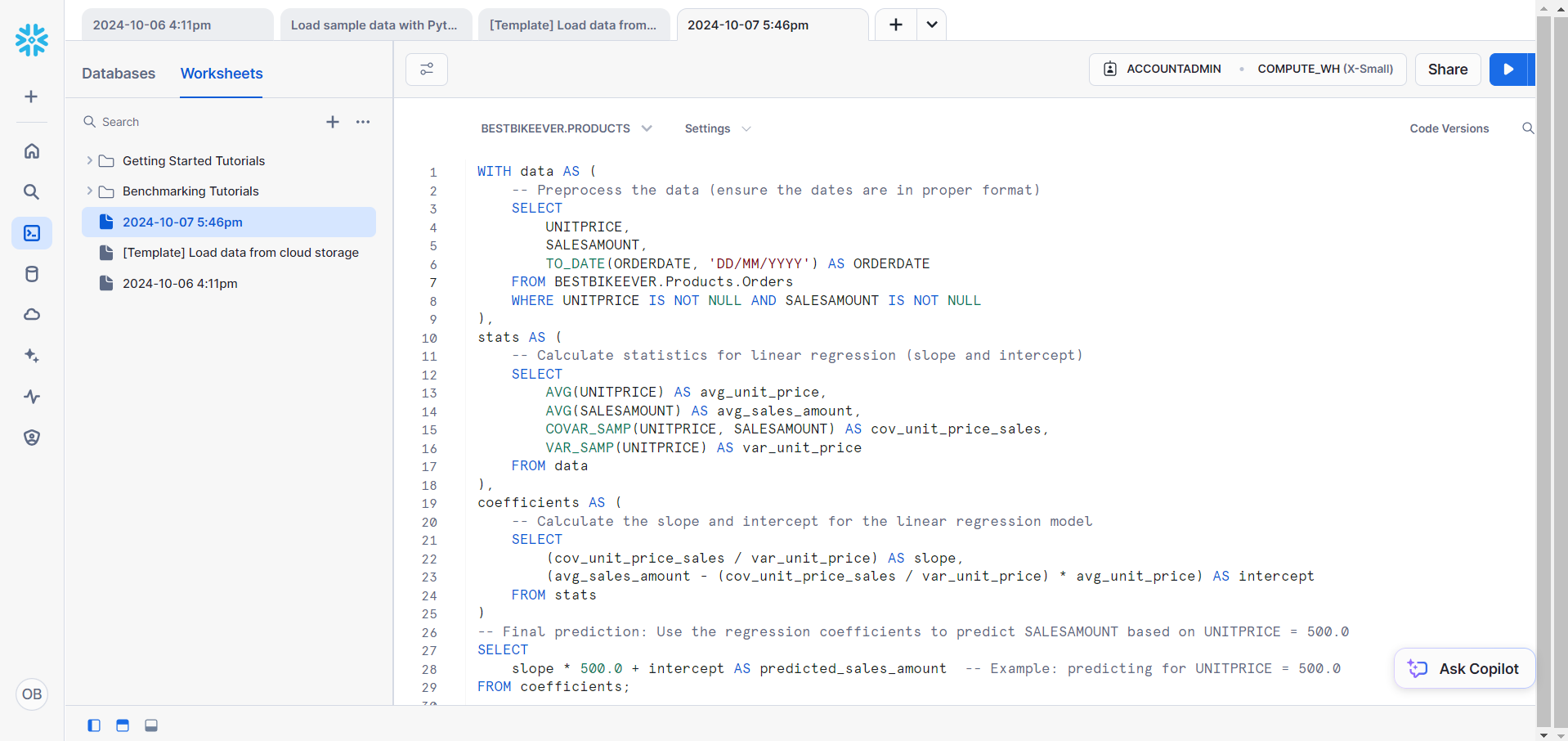
Password: DiwAli@#2023

# Predictive Modelling in Snowflake

We developed a linear regression model in Snowflake which can be seen in figure 1 to predict sales based on key features such as UNITPRICE and SALESAMOUNT. To enhance accuracy, additional variables like PRODUCTSTANDARDCOST were included in the model.

The model was created directly in Snowflake using SQL, without needing external tools for basic predictive tasks. We calculated the regression coefficients (slope and intercept), which provided a simple yet effective model for forecasting sales. The predictive query demonstrated how UNITPRICE could be used to forecast SALESAMOUNT, enabling BBE to predict future sales based on historical trends.

Integration with External Tools:  
Snowflake supports live connections to Tableau and Jupyter for more advanced workflows. Through Tableau, the data can be visualized in interactive, real-time dashboards. For more complex predictive models and machine learning, Snowflake's Snowpark for Python as depicted in figure 2 allows integration with Jupyter Notebooks, enabling Python-based analysis within Snowflake.



[Link to predictive query](https://app.snowflake.com/dfubbbq/kwb90249/w45MQWomXDhu/query)

Fig1: Predictive modelling in Snowflake

# A screenshot of a computer Description automatically generated

Fig 2: Connecting Snowflake with Jupyter Notebook via snowspark

# Expanded Section on Dashboard Creation in Snowflake and Tableau

## Dashboard Creation in Snowflake

In Snowflake, we leveraged its robust SQL capabilities to create a data-driven sales dashboard as shown in figure 3, this dashboard provides valuable insights into sales performance and profitability and focusses on key metrics that allow decision-makers to track and predict trends effectively.

## Key Graphs in Snowflake:

### Sales Trends:

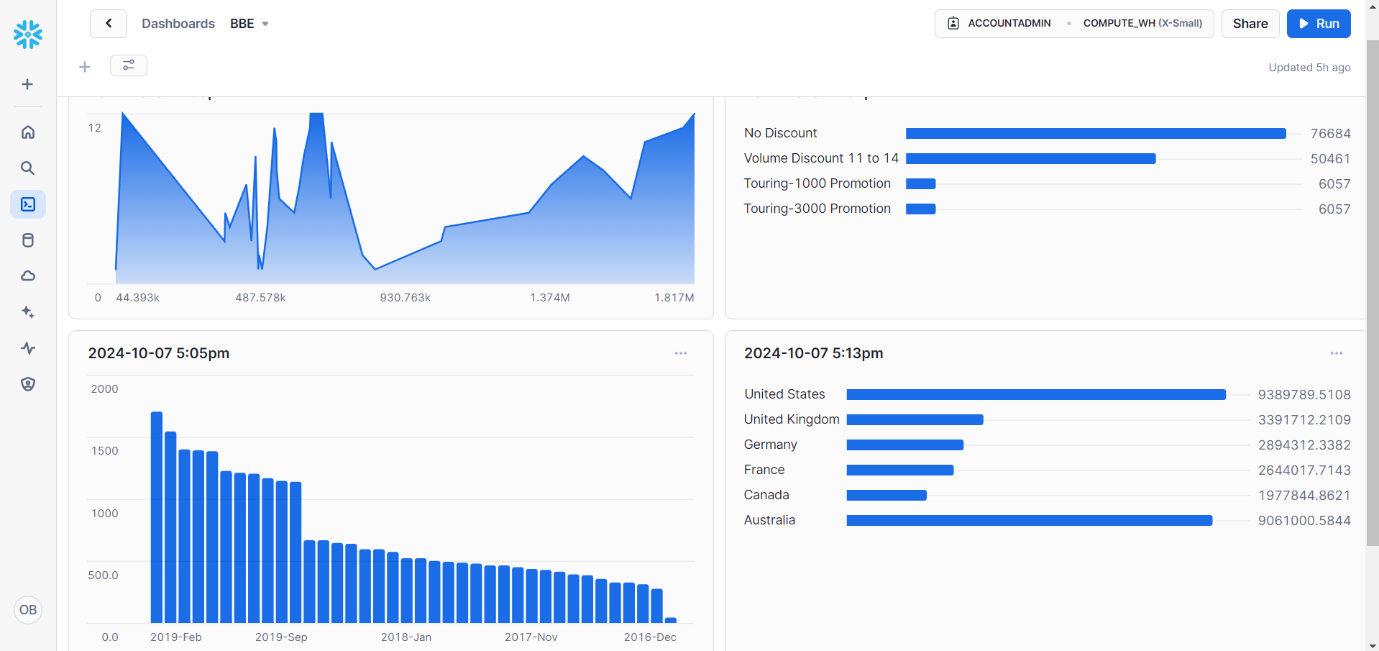
This graph captures historical sales data, enabling BBE to identify trends over time. Using the predictive linear regression model, the graph forecasts future sales, offering a comprehensive view of sales progression. The **sharp peaks and troughs** illustrate periods of high and low sales activity, which can help BBE identify seasonal fluctuations or market disruptions.

### Promotion Impact on Sales:

This bar chart highlights how various promotions influence sales performance. For example, the graph demonstrates that sales with "No Discount" outperform other promotions significantly. Meanwhile, volume discounts and specific promotions like "Touring-1000" and "Touring-3000" contributed to sales but with lower figures. These insights are critical for BBE to strategize their future promotions.

### Sales by Region:

A horizontal bar chart breaks down sales by geographic regions such as the **United States**, **Australia**, and **Europe**. The graph shows that the **United States** and **Australia** have the highest sales figures, while regions like **Germany** and **Canada** contribute less to comparison. This regional breakdown helps BBE identify markets where they are performing well and regions that need more attention.



[Snowflake Dashboard](https://app.snowflake.com/dfubbbq/kwb90249/#/bbe-dE6xLQHIL)

Fig 3: Dashboard using Snowflake

While Snowflake's dashboards are powerful, connecting to **Tableau** allows for even more interactive and detailed analysis. By establishing a **live connection** from Snowflake to Tableau, the data remains up-to-date, providing real-time insights without needing manual updates.

## Key Graphs in Tableau:

### Sales by Region (Map Visualization):

This interactive map shows a geographical breakdown of sales, making it easier for BBE to visualize where their products are most popular. Regions like **North America** and **Europe** are highlighted, with deeper colours indicating higher sales. This visualization is particularly useful for spotting regional trends and expanding into new markets.

### Profit Contribution by Product:

This line graph tracks the **profit contribution** of different products over time, with a forecast line showing predicted profit contributions. The graph provides insights into how much revenue each product generates, allowing BBE to prioritize high-performing products and identify underperforming ones.

### Sales Forecast:

This graph predicts future sales based on historical data, displaying both actual and estimated sales figures. The **forecast line** helps BBE anticipate potential sales figures, guiding decision-makers to prepare for both growth and challenges ahead.

### Effect of Yearly Income and Commute Distance on Sales:

This scatter plot analyzes the impact of customer demographics on sales. Specifically, it looks at how **yearly income** and **commute distance** influence **unit prices**. This data helps BBE tailor their marketing efforts toward specific customer segments, ensuring more targeted advertising.

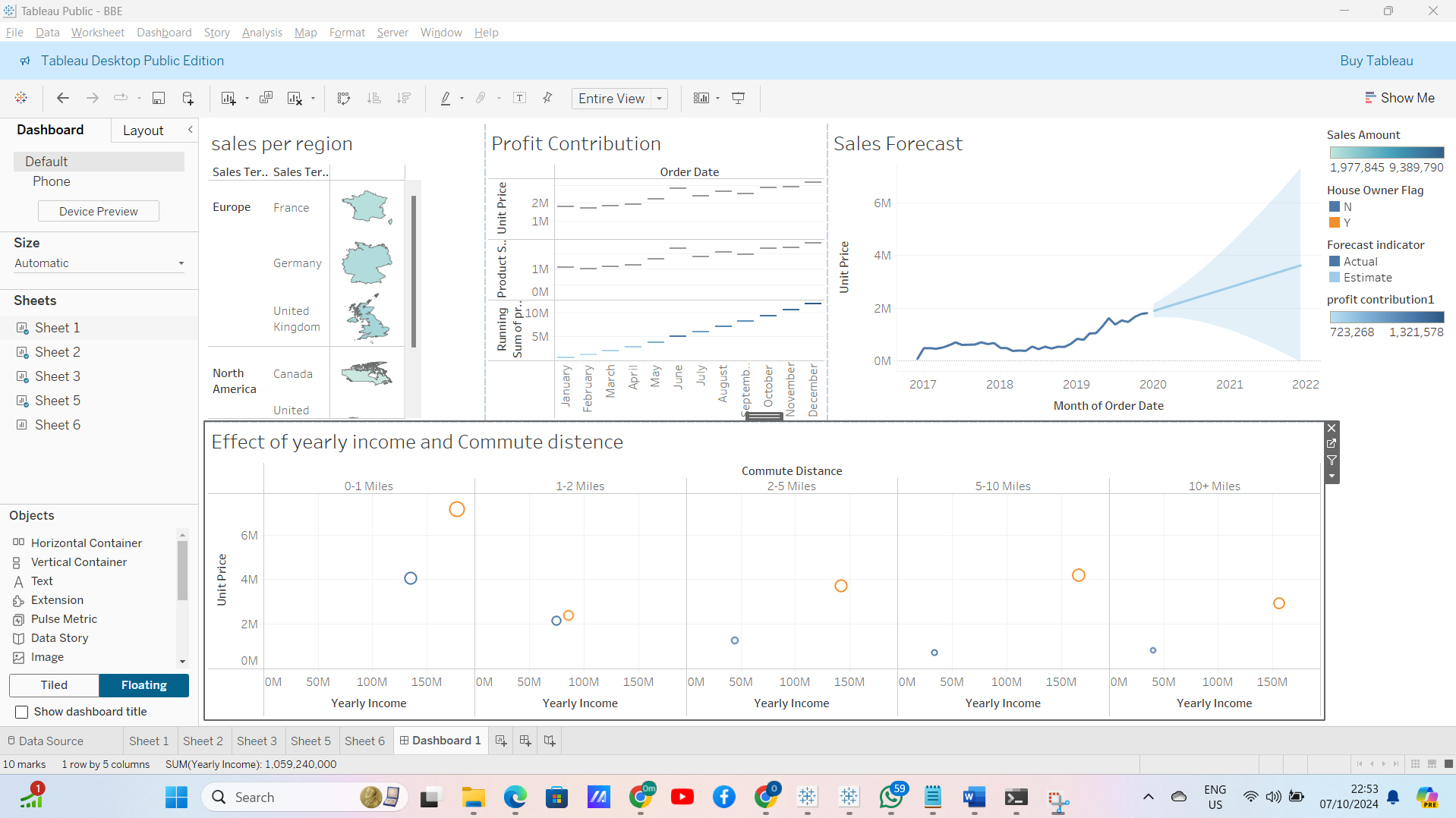


Fig 4: Tableau dashboard connected to Snowflake for real-time visualizations.

# Conclusion:

Both the Snowflake-generated dashboard and the Tableau-enhanced visualizations offer BBE critical insights into their sales performance and product profitability. Snowflake's live connection capabilities with Tableau as displayed in fig 4 above, ensures that the data is always up-to-date, providing a powerful tool for making informed, data-driven decisions.

## Comparison: Microsoft Azure Synapse Analytics, Snowflake, and Google BigQuery

When comparing Microsoft Azure Synapse Analytics, Snowflake, and Google BigQuery, all three are top-tier data warehousing platforms, each with its own strengths in scalability, performance, and integration. Here’s a more user-friendly comparison highlighting their key features:

### 1. Scalability and Performance

**Snowflake**: Known for its seamless scalability, Snowflake separates compute and storage, making it easy to scale resources as needed without any downtime. Its automated performance tuning ensures your queries run efficiently, even with large datasets.

**Azure Synapse**: Combines big data processing with traditional data warehousing, offering both dedicated and serverless SQL pools. It's a powerful option for those already using the Azure ecosystem but may require more manual effort to optimize performance.

**Google BigQuery:** With its serverless design, BigQuery automatically scales to meet your needs. It performs exceptionally well with large queries, thanks to its columnar storage but may need fine-tuning for very complex tasks.

### 2. Ease of Use

**Snowflake**: Designed to be user-friendly with minimal setup, Snowflake automates many processes, making it accessible even for non-technical users. Its easy integration with tools like Tableau and Jupyter simplifies advanced analytics and visualization.

**Azure Synapse**: While it offers a lot of flexibility, Azure Synapse can be more complex to set up and manage, particularly if you're not already familiar with the Azure environment.

**Google BigQuery**: It's straightforward, especially for those already using Google Cloud, but the query-based pricing model can be tricky to estimate for users running frequent or large queries.

### 3. Integration and Ecosystem

**Snowflake**: Plays well with others, offering broad integration with AWS, Azure, and Google Cloud, as well as a variety of third-party business intelligence (BI) tools. Its flexibility means you're not locked into any one provider.

**Azure Synapse:** Best suited for those deeply integrated with the Microsoft ecosystem, with excellent connections to Power BI, Azure ML, and Azure Data Lake.

**Google BigQuery:** Works seamlessly within the Google Cloud environment, with strong connections to tools like Google Data Studio and Looker.

### 4. Cost Efficiency

**Snowflake**: Operates on a pay-as-you-go model, charging separately for compute and storage. It also has an auto-suspend feature to save costs when resources aren’t being used.

**Azure Synapse:** Offers both provisioned and serverless options. While flexible, mismanagement of resources can lead to higher costs, especially in provisioned mode.

**Google BigQuery:** Uses a query-based pricing structure, which is great for smaller or infrequent queries but can become costly with heavy usage.

### Conclusion:

**Snowflake** is the most well-rounded option, offering simplicity, scalability, and flexibility across cloud providers. Azure Synapse excels for businesses already deeply invested in the MicrosoftAzure ecosystem. Google BigQuery is ideal for GoogleCloud users, offering high performance with a serverless model.

In summary, Snowflake stands out as the go-to platform for businesses looking for an easy-to-use, flexible, and cost-effective data warehousing solution.