**Name:** Obinna Williams

**Student ID:** VEPH/20B/DA183

TASK 16B

**TECHNICAL REPORT FOR NORTHSTAR METRICS: BUSINESS INSIGHTS AND STRATEGIC ANALYSIS FOR THE YEAR 2019**

**Outline**

* **Introduction**
* **Story of Data**
* **Data Splitting and Pre-processing**
* **Pre-Analysis**
* **In-Analysis**
* **Post-Analysis and Insights**
* **Data Visualization & Charts**
* **Recommendations and Observations**
* **Conclusion**
* **References & Appendices**

**Introduction**

1. **Objective of the Project**

The primary objective of this data analysis is to provide insights into business performance and sales trends for the year 2019. It helps in identifying top performing salespersons, best-selling products, key customers, and high-revenue cities.

1. **Problem Being Addressed**

The analysis aims to answer key business questions, such as:

* Which products and regions generate the highest sales?
* Who are the top-performing salespersons?
* What are the sales trends over the months?
* Which customers contribute the most to revenue?
* How are sales distributed across different regions and cities?

1. **Key Datasets and Methodologies**

Datasets Used:

* Sales Data (Revenue, Transactions, Product Categories, Regions)
* Customer Data
* Salesperson Performance Data
* Regional and City-Based Sales Data

Methodologies in Excel:

* **Pivot tables & Pivot charts** to summarize sales performance by product, region, and salesperson.
* **Bar charts & Pie charts** to visualize top-performing customers, sales trends, and revenue distribution.
* **Line Graphs** to track monthly sales trends.
* **Filters & Slicers** to segment data for detailed insights.

**Story of Data**

1. **Data source**

This dataset is sourced from Kaggle.com

1. **Data Collection Process**

The data collection process is by web scraping.

1. **Data Structure**

The dataset is organized in a tabular format, where:

* Each row represents a sales transaction.
* Each column represents a variable, detailing aspects like customer information, shipping details, payment type, and sales revenue.

1. **Important features and their significance**

**Order Details:**

* Order ID: A unique identifier foreach order.
* Order Date: When the order was placed, useful for trend analysis.
* Customer ID & Customer Name: identifies repeat buyers and top customers.

**Customer & Sales Information:**

* Salesperson: Helps track individual sales performance.
* Region: Allows for regional sales comparison.
* Product Name & Category: Determines best-selling products.

**Financial Metrics:**

* Unit price, Quantity, Revenue: essential for calculating total sales and profitability.
* Payment Type: Indicates preferred payment methods.

**Shipping & logistics:**

* Shipped Date & Shipper Name: Tracks delivery times and carrier performance.
* Shipping Fee: Helps analyse logistics costs.

1. **Data Limitations or Biases**

While the dataset provides valuable insights, there are potential limitations that impacts the accuracy and interpretation of the analysis. Below are some key Issues:

Missing or Incomplete Data:

* Order Gaps
* Customer Details
* Shipped Date

**Data Splitting and Preprocessing**

1. **Data Cleaning**

To ensure accuracy and consistency, the following steps were taken to clean the data:

* Removing Duplicates: Identified and eliminated duplicate entries using Excel’s built in functions.
* Correcting Errors: Checked for inconsistencies in product names, customer details, and region fields.
* Handling Formatting Issues: Standardized data formats and ensured numerical values were correctly formatted.

1. **Handling Missing Values**

* There were no missing values.

1. **Data Transformation**
2. **Data Splitting**

Independent Variables: These are factors that influence sales but don’t rely on other data points,

* Customer Name.
* Salesperson.
* Region.
* Ship Name.
* Ship Country.
* Payment type.
* Category.
* Address.
* Order date.
* Product Name.

Dependent Variables: These are factors that change based on independent variables,

* Revenue.
* Unit price.
* Shipping Fee.
* Order Fee.
* Quantity Sold.

1. **Industry Context**

* Industry Type: A Grocery store with revenue means.
* Relevance: The dataset reflects sales performance, customer behaviour, and supply chain efficiency within the grocery store.

1. **Key Stakeholders**

* The Chief Executives: To utilize insights for revenue optimization and strategic planning.

1. **Value to the Industry**

* Generating more revenue. This is what is most important to them.

**Pre-Analysis**

**Potential Analysis:**

* Best performing region by revenue generated.
* Performance analysis of each salesperson by revenue.
* Customer's performance analysis base on revenue.
* Customer's performance analysis base on quantity purchased.
* Payment type usage.
* Best performing goods for the year by quantity.
* Best performing goods for the year by revenue.
* Performance analysis for the shipping countries.

**Potential Insights**

* Finetune the best region and explore more measures to gain more grounds against competitors e.g. more advertisements, we can also finetune the least performance region and implement measures done in the best performance region and ensure these measures are kept in place by transferring the sales person in the region with great sales to the region with low sales to ensure that they implement measures enacted at the best performing regions.
* Most used Payment type.
* Stocking most purchased goods.
* The country we shipped to the most

**In-Analysis**

1. **Unconfirmed Insights:**

* December has the highest revenue at $66,642.78, suggesting a peak sales period, likely due to holiday shopping or year-end demand.
* February has the lowest revenue at $19,955.50, which may indicate a post-holiday slowdown.
* Company D has the highest revenue at $67,180.50, contributing significantly to the overall total.
* Nancy Free Hafer has the highest revenue at $104,242.34.
* Jan Kotas has the lowest revenue at $16,350.50.
* The north region generated the highest revenue at $141,660.34.
* The west region generated the lowest revenue at $91,251.98.
* Beverages like coffee, Beer, Chai have strong sales and shows high demand.
* Most transactions 218, fall in the $0-$1000 range, indicating that most sales are of lower value.
* New York generates the highest revenue at $67,180.50, making it both the best performing city and shipping city

1. Preliminary Recommendations:

* Sales peak in December, suggesting that marketing efforts and promotions should align with holiday shopping trends.
* Months like February, April, and July that have the lowest revenue, strategies like discounts or new product launches can help drive sales during these months.
* Losing any of the top three companies could significantly impact revenue, making customer retention strategies essential.
* Jan has generated the lowest revenue and i suggest that she may require additional sales and customer retention training.
* I recommend that the top three performers should be retained and rewarded with incentives like bonuses or promotion.
* Sales strategies implemented the best region (North) should also be implemented to the least performing regions to boost sales in those regions.
* Beverages like coffee, beer chai show high sales, i suggest an expansion in beverage offerings
* Most revenue are generated from frequent small value transactions, indicating that the business may rely on volume sales rather than a few high value sales.
* New York remains an important revenue hub, so maintaining a strong distribution and customer engagement there is vital.

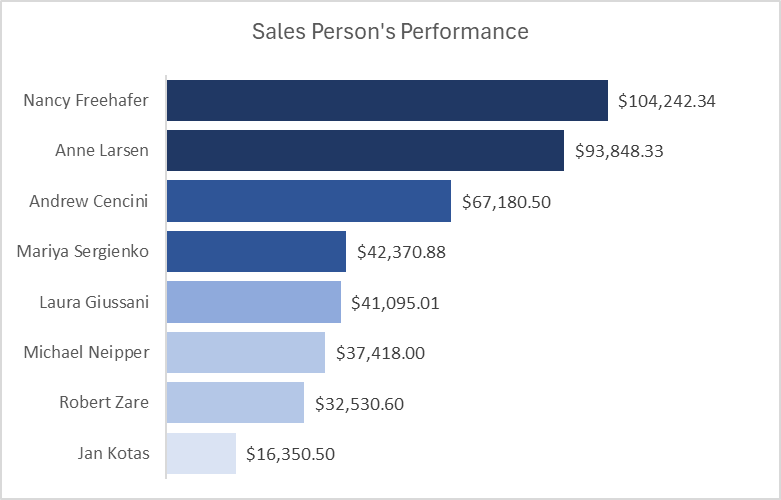
1. **Analysis Technique used in Excel:**

* Pivot Tables and Pivot Charts

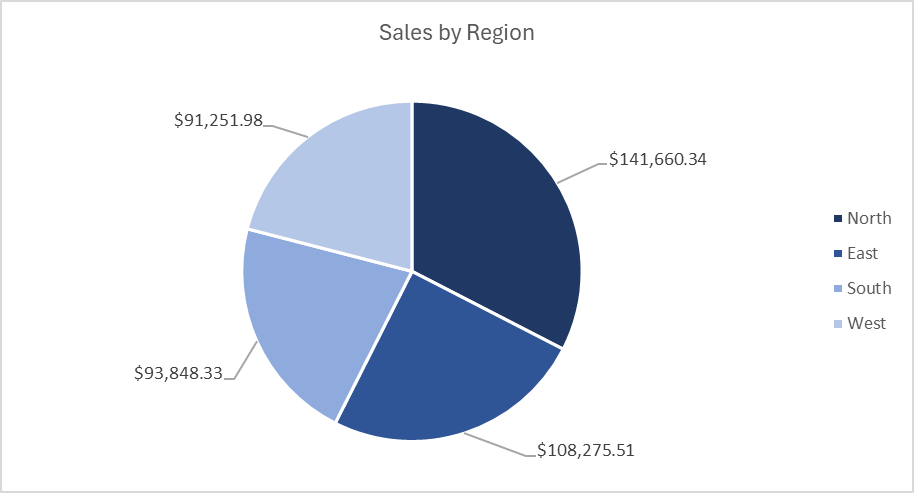
**Data Visualization & Charts**

1. **Charts and Graphs:**

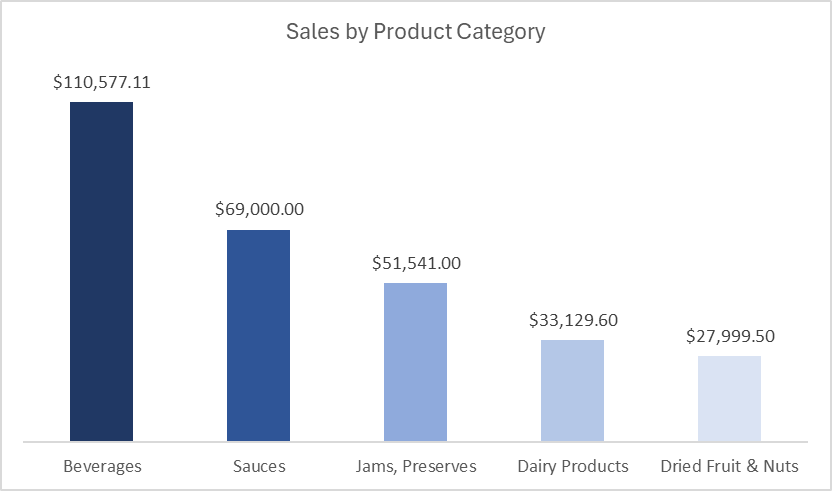
* The Line graph shows a peak in May at $55,601.61, followed by a sharp drop to $27,318.34 in June.
* The bar chart shows Customer D as the top customer with the highest sales at $67,180.50

****

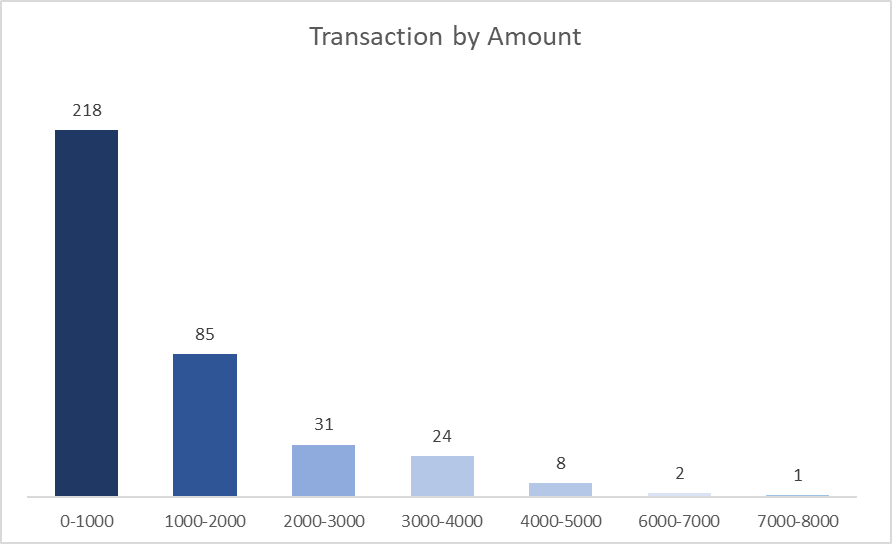
* The bar chart shows Nancy Freehafer leads with the highest sales at $104,242.34

****

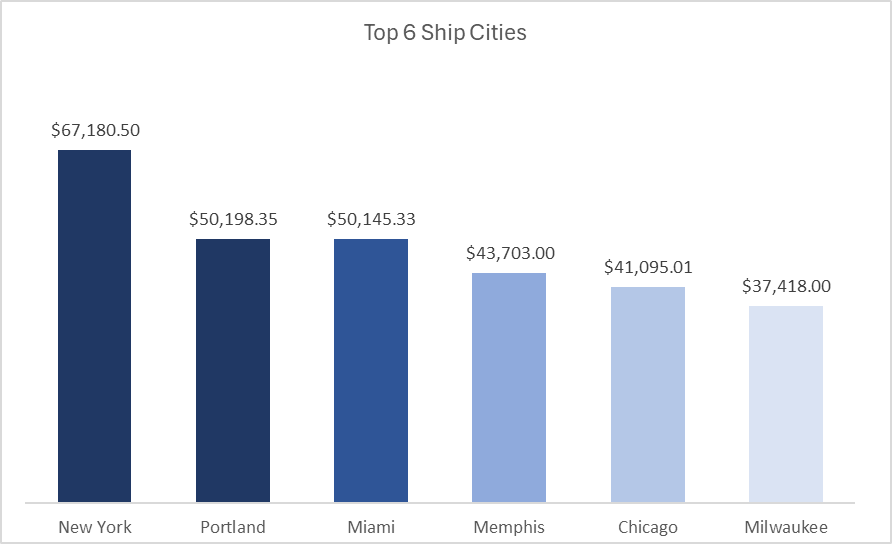
* The pie chart shows the North region has the highest sales at $141,660.34, making up the largest portion of the pie.

****

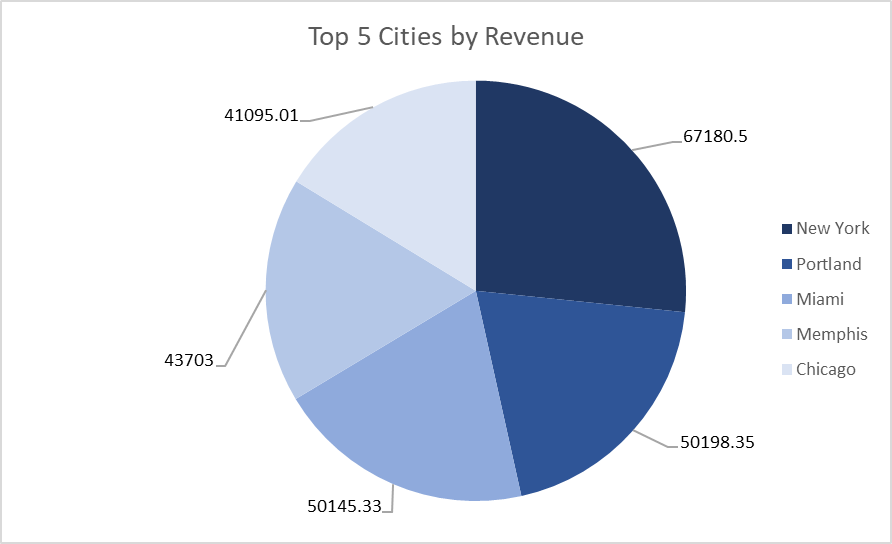
* The column chart shows Beverages lead with the highest sales at $110,577.11, depicted by the tallest bar.

****

* The bar chart shows the 0-1000 range has the highest number of transactions at 218, depicted by the tallest bar.

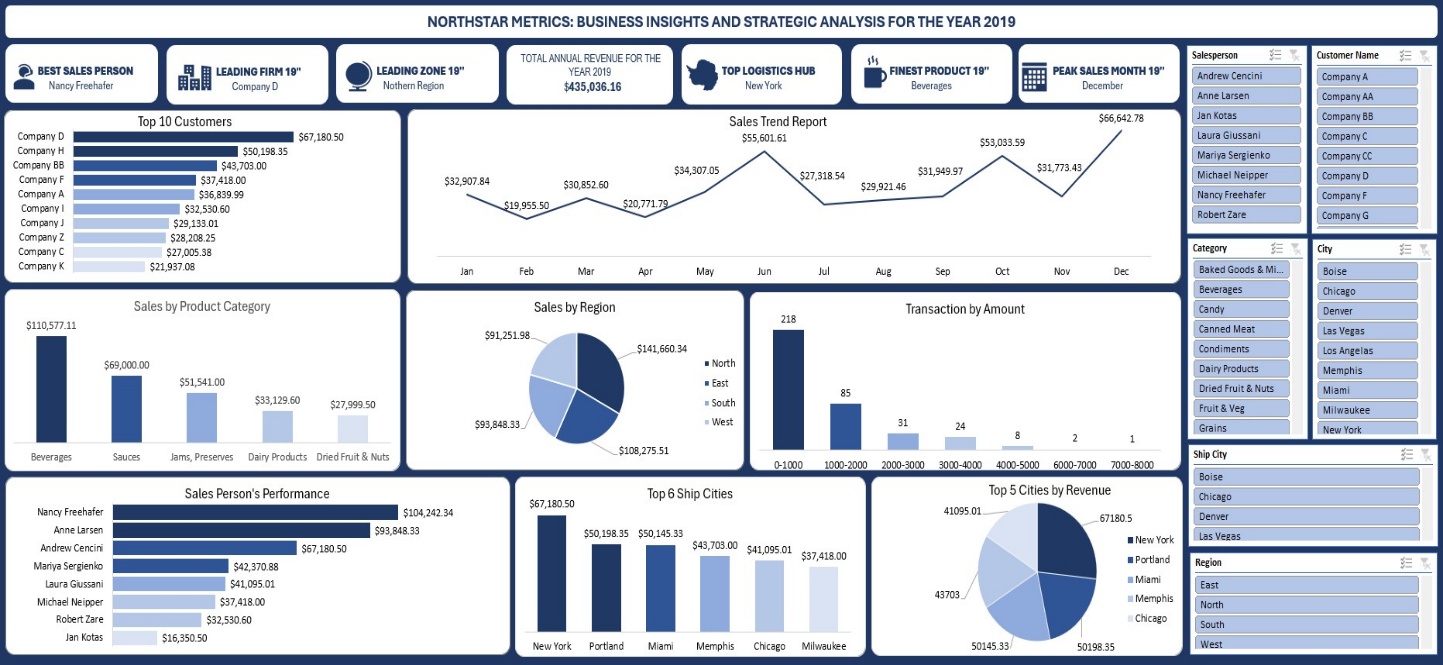
****

* The bar chart shows New York leads with the highest value at $67,180.50.

****

* The pie chart shows that New York has the highest revenue at $67,180.50, making up the largest portion of the pie.

1. **Dashboard:**



**Recommendations and Observations**

1. **Observations:**

**Salesperson:**

**ANNE LARSEN**

1. Anne Larson generated a total revenue of $50,145.33 and $43,703.00 from Miami and Memphis.

2. Anne is responsible for a significant share of sales from company BB, company k, and company z.

3. Anne's best-selling product category is beverages with a revenue of $29,278.08

4. Anne's sales are primarily from Miami, and she is a major contributor to the south Region's performance.

5. From the sales trend report chart, Anne's best sale month is June with a revenue generation of $10,811.70

**JAN KOTAS**

1. Jan Kotas generated a total revenue of $16,350.50 from Denver.

2. Jan Kotas generated the highest revenue of $4,928 in the month of June.

3. Jan Kotas highest product sales are the beverages and fruit & Vegetables with a revenue generation of $6,986.00 and $ $6,942.00

1. **Recommendations:**

**Salesperson:**

**ANDREW CENCINI**

1. Since all sales of Andrew comes from company D, I recommend that Andrew work on acquiring more clients to reduce dependency risk.

2. Andrew's sales are focused on only new work and the northern region, i suggest that Andrew expands into other high performing locations like Chicago and Miami.

3. Most of Andrew's transactions fall below $1000, which indicates a high volume of small deals, i suggest Andrew targets bigger clients so he can boost revenue with fewer transactions.

4. I recommend that Andrew should capitalize on high sales period by running targeted campaigns and offering promotions.

**ANNE LARSEN**

1. Since Anne generated a significant portion of her revenue from Miami, she should focus on strengthening her customer base in this city by reaching out to more clients and offering exclusive deals.

2. In June Anne had her best sales, Anne should analyse what contributed to that success and implement similar strategies to maintain strong sales throughout the year.

3. Anne's best product category is beverages, but relying too much on a single product category can be risky. she should explore increasing her sales in other product categories, such as Soup, to ensure a balanced portfolio.

4. Anne should promote bulk sales for beverages since this is her best performing product category, and Anne should provide discounts on large orders.

**Conclusion**

1. **Key Learnings:**

* Seasonal trends and regional performance: December is the peak sales month ($66,648.78), likely driven by holiday demand, while February is the lowest ($19,955.50). The North region leads in sales ($141,660.34), indicating a strong market presence there.

1. **Limitations:**

* The analysis covers only one year, limiting the ability to identify long term trends or seasonality patterns beyond the observed year.

1. **Future Research:**

* Analyse multiple years of data to confirm seasonal patterns and identify long-term trends, especially for underperforming months like February.

**References & Appendices**

1. **Data Source:**

* Kaggle.com

1. **Tools:**

* Microsoft Excel.
* Pivot Tables and Pivot Charts.