**Diagnostic Analytics Report – Retail Sales Analysis Using Microsoft Azure**

# Executive Summary

# This report is a diagnosis analytic project to find the main causes of reduction in sales of a retail firm. The project takes advantage of the cloud-based Microsoft Azure ecosystem, namely Azure Data Factory, Azure Synapse Analytics, and Power BI, to illustrate how the latest generation of data analytics solutions may be utilized to derive actionable insights with regard to raw data and inform business strategy. This analysis was based on a publicly available dataset (Superstore Sales Dataset). Azure Data Factory was used to import and clean the data, and then Azure Synapse Analytics was used to explore the data using SQL to determine the trends in sales, variations in performance by regions, and customer behavior. Then, Power BI dashboards were created to visualize these results to gain more insights about trends over time, space, and product types. It was found that sales vary greatly with season, especially high in June and December, and that some regions, like the South were underperforming. Also, there were some sub-categories of products, such as Tables and Bookcases, where the profits were constantly negative because of over-discounting. On the other hand, other categories like Technology and Office Supplies were doing good in sales as well as profitability. As part of the derived insights, the report provides a number of data-driven recommendations such as the optimization of discount policy, shifting marketing resources to the regions with high potential, concentration on the profitable product lines, and improvement of customer retention campaigns by targeting them. Comprehensively, the given project demonstrates the usefulness of diagnostic analytics in retailing, which ultimately confirms how cloud-based systems, such as Microsoft Azure, can enable an enterprise to make competent decisions, enhance the effectiveness of operations, and be proactive with regard to the fluctuating market trends and conditions.

# Objective

# The main purpose of the project is to create a diagnostic analytics solution with the help of Microsoft Azure in order to explore and provide insights into the reasons of reduction in sales within a retail company. Through organized analytical thinking and the use of cloud-based tools, this project will seek to determine patterns, correlations and anomalies in the sales data that could be used to make strategic decisions.

# Key goals include:

# Preparing a real-world retail dataset in Azure by importing it with the help of Azure Data Factory.

# Analyzing the trends of sales performance using SQL-based analysis in Azure Synapse Analytics.

# Using visualizations as a way of discovering insights through interactive dashboards in Power BI to show correlation between sales, regions, product category and customer behavior.

# Determining factors that contributed to the witnessed reduction in sale within designated segments or geographies.

# Providing actionable, data-driven recommendations to improve retail performance and align business strategies with evidence-based insights.

# Azure Setup

I created a Microsoft Azure account and then explored services such as Azure Data Factory, Synapse Analytics, and Azure Data Lake Storage. Below is the evidence of the Azure portal setup:

A screenshot of a computer

Description automatically generated

**The Azure Portal served as the centralized hub for managing all services used in the diagnostic analytics project. Key resources included:**

* **Azure Synapse Workspace (conestogaworkspacenitigya) – for running SQL queries and analytics.**
* **Azure Storage Account (conestoganitigyastorage) – for storing the raw Superstore dataset in blob format.**
* **Azure Data Factory (R1NV) – to build ETL pipelines for data import and transformation.**
* **Resource Group (R1\_NV) – to organize related services for streamlined management**

# Data Preparation and Cleaning

1. **Dataset Selection**

* **Dataset Used:** *Sample - Superstore Sales Dataset* (CSV format).
* **Source:** Kaggle (publicly available).
* **Reason for Selection:** Includes transactional sales data across categories, regions, and customer segments.

1. **Azure Blob Storage Configuration**

* Created a new **Storage Account** in the Azure portal.
* Set up a **Blob container** named superstore-data.
* Uploaded the CSV file from the local system into the container.

1. **Azure Data Factory – Pipeline Setup**

* Created a new instance of **Azure Data Factory**.
* Launched the **Copy Data tool** to initiate the import pipeline:
  + **Source:** Azure Blob Storage (pointing to the Superstore CSV).
  + **Sink/Destination:** Azure SQL Database configured via Linked Service.

1. **Sampling and Aggregation within ADF**

Before sending the data to the destination (sink), a **transformation step** was performed using **Data Flow** in ADF:

* **Data Flow Configuration**:
  + Added a new **data flow activity** to the pipeline.
  + Applied **aggregate transformation** to group and summarize the data.
* **Aggregation Logic**:
  + Grouped data by **Category** and **Region**.
  + Calculated:
    - Total Sales (SUM(Sales))
    - Average Discount (AVG(Discount))
    - Total Profit (SUM(Profit))
  + The result was a summarized table representing business performance across each category and region.
* **Validation**:
  + Used the Data Flow Debug mode to preview the output.
  + Ensured all numerical columns were aggregated properly.

1. **Connecting to Azure Synapse Analytics**

Created a new **Azure Synapse Workspace**.

Added a **dedicated SQL pool**.

Connected to the database where the cleaned and aggregated data was stored.

Queried the data using Synapse Studio to validate the aggregation and prepare for deeper diagnostics.

**SQL Queries and Analytical Insights**

1. Monthly Sales Trend AnalysisA screenshot of a computer

Description automatically generated

**Insight**:

* Clear seasonality is observed with **peak sales in June and December** every year.
* **Mid-year months (August–October)** consistently underperform.
* The company can use this trend for **campaign planning, inventory stocking**, and **seasonal promotions**.

2. Region-wise Total SalesA screenshot of a computer

Description automatically generated

**Insight**:

1. The **West region** leads in overall sales, followed by **East** and **Central**.
2. The **South region** underperforms significantly—indicating a need for targeted marketing or local analysis.

Top 10 Customers by Total Sales A screenshot of a computer

Description automatically generated

**Insight**:

* Customers like **Sean Miller**, **Tamara Chand**, and others significantly contribute to revenue.
* This opens opportunities for **personalized loyalty programs**, **VIP perks**, or **exclusive offers**.

4. Category-wise Total Profit A screenshot of a computer

Description automatically generated

**Insight**:

* **Technology** and **Office Supplies** are the most profitable categories.
* **Furniture**, despite strong sales, yields **lower profit** due to **higher return rates or discounts**.
* Reassessing product margins or pricing in the Furniture category could enhance overall profitability.

# Power BI Dashboards and Visualizations

# 1. Total Sales by Year :

# A blue graph with white text Description automatically generated

#### ****Insight:****

* **The line and area chart shows a clear rise from 2014 to 2015, reaching the highest sales volume in that year.**
* **From 2015 onward, a sharp and consistent decline is visible through 2016 and especially 2017.**
* **The fall in 2017 suggests a significant drop in customer engagement or operational efficiency.**
* **This pattern reflects a short-lived growth spike without long-term reinforcement.**

#### ****Possible Reason:****

**The company likely capitalized on a one-time sales peak in 2015 without retaining those customers or diversifying its channels, resulting in unsustainable growth and eventual decline.**

**2. Total Sales by Category and Year :**

**A graph of blue squares

Description automatically generated**

#### ****Insight:****

* **Furniture sales skyrocketed in 2015**, dominating that year’s revenue share.
* In 2016, **Technology took the lead**, showing a shift in customer demand or sales focus.
* **Office Supplies remained steady** but never dominant.
* All three categories collapsed in 2017, showing that **no category had long-term resilience or brand loyalty**.
* The category shifts suggest frequent **strategic redirection** rather than building consistent product demand.

#### ****Possible Reason:****

**Frequent changes in product promotion priorities likely confused the customer base and fragmented marketing impact, preventing any category from establishing long-term traction.**

**3 Total Sales by Region and Year :**

**A graph of blue bars

Description automatically generated with medium confidence**

#### ****Insight:****

* The **West region led in 2014** but saw a continuous and dramatic decline by 2017.
* **Central and East** showed strong performance in 2015 and 2016 but couldn’t sustain growth.
* **South** remained a **low contributor** every year, suggesting under-penetration.
* **2017 marks a synchronized drop across all regions**, hinting at system-wide failure rather than isolated geographic challenges.

#### ****Possible Reason:****

**Failure to implement regionalized strategies (e.g., localized promotions, tailored product offerings, regional partnerships) caused engagement fatigue and left high-potential markets underdeveloped.**

**4. Monthly Total Sales Comparison by Year :**

# A screenshot of a graph Description automatically generated

### 1. ****September & June Spikes****

* **2015 (dark blue)** shows **massive spikes** in **September** and **April**, suggesting strong seasonal campaigns or bulk orders during those months.
* **2014 (light blue)** had a big surge in **June**, possibly due to summer demand or promotional events.

**Possible Reason**: End-of-quarter discounts or financial-year closing purchases in businesses.

**2. December & November Peaks**

* **2016 (orange)** peaks in **December**, followed by **November**.
* Suggests **holiday season sales** (Black Friday, Christmas, etc.) were strongest in 2016.

**Possible Reason**: More effective seasonal campaigns, or wider product offerings during holidays.

**3. Consistently Low Performance in 2017**

* **2017 (purple)** shows **low sales across all months**, especially compared to previous years.
* December barely breaks 200 units.

**Possible Reason**: Budget cuts, fewer campaigns, or data availability issues (partial data).

**4. Flat Months Across Years**

* **January, March, May, July** all have generally **low sales** for every year.
* Indicates **seasonal low** or off-peak retail periods.

**Insight**: These are likely **planning or inventory months**, not peak sales windows.

**Regional Sales Contribution by Year (2014–2017) :**

A graph of different colored squares

Description automatically generated

**1. 2015 – Peak Year with Broad Regional Support**

# Total sales peaked in 2015.

# East (orange) and South (purple) contributed significantly.

# West (dark blue) still strong but less dominant than in 2014.

# Insight: Likely due to company expansion or regional campaigns hitting their stride in East & South.

# 2. 2014 – West Dominance

# West region (dark blue) contributed the majority of sales.

# Central (sky blue) and others contributed far less.

# Insight: Sales likely concentrated in established West markets during early operations.

# 3. 2016 – Decline Begins

# Overall sales dropped compared to 2015.

# Contributions from all regions flattened, especially West.

# Insight: Possible market saturation, reduced promotions, or economic factors.

# 4. 2017 – Lowest Year for All Regions

# All regional sales dipped sharply.

# Even strongholds like Central and East show minimal contribution.

# Insight: Could be a partial dataset or major operational shift (e.g., fewer SKUs, data truncation, business slowdown).

# Monthly Seasonal Sales Trends by Year :

# A graph of different colored columns Description automatically generated

**Key Insights from the Chart:**

**1. Seasonal Peaks Observed in Specific Months**

* **June**, **September**, and **December** consistently show **higher total sales**, highlighting **seasonal shopping peaks** likely driven by mid-year sales, back-to-school shopping, and holiday season promotions.
* **December** shows strong year-over-year contributions, especially from **2016**, which dominated sales during that month.

**2. Dominance of Specific Years in High-Sales Months**

* **2014** had the highest sales in **June**, showing a significant spike that other years didn’t match. This could indicate a successful mid-year campaign or business growth phase.
* **2015** had major contributions in **September** and **April**, suggesting stronger performance in the second and early quarters of the year.
* **2016** dominated **December**, outperforming all other years—potentially due to strategic promotions or market expansion.
* **2017** had the **lowest contributions overall**, indicating a decline in performance or incomplete data for that year.

**3. Low Performing Months Across All Years**

* **January**, **February**, and **March** have consistently low sales volumes. This may be due to post-holiday spending fatigue or limited marketing focus early in the year.

**4. Sales Volatility and Lack of Uniform Seasonal Trends**

* While some months show recurring trends (e.g., December spikes), others like **April** and **August** display erratic year-to-year changes.
* This suggests either inconsistent campaign planning or external factors like economic fluctuations or product lifecycle changes.

**5. Yearly Trend Visibility Through Stacks**

* By stacking years within each month, we can see:
  + Which years outperformed or underperformed month-by-month.
  + If a month’s sales strength is due to **one strong year** (e.g., 2014 in June) or **consistently moderate performance across years**.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Metric | 2014 | 2015 | 2016 | 2017 | Key Insight |
| Total Sales (USD) | $45,273 | $56,329 | $61,782 | $59,614 |  |
| Top Month | December | December | September | December | December consistently strong for promotions. |
| Lowest Month | January | February | January | March | Q1 underperforms, needs strategic focus. |
| Top Region | West | East | West | West | West dominates; East spikes in 2015. |
| Top Segment | Consumer | Consumer | Consumer | Consumer | Consumer is the most profitable segment. |
| Top Category | Office Supplies | Furniture | Office Supplies | |  | | --- | | Technology | |  | | Indicates shifting yearly demand. |

# Actor Identification and Interaction Diagram

1. **Customer (Data Source)**

* Represents the origin of raw data (e.g., retail systems or export files).
* Pulls and pushes raw transactional data (sales, customer, order details) into Azure Data Lake.

1. **Azure Data Lake**

* Serves as the centralized cloud repository for storing raw, unstructured, or semi-structured data.
* Stores large volumes of raw data uploaded by the Customer for processing.
* Acts as the starting point for downstream data transformations.

1. **Azure Data Factory**

* The Data Engineer uses this to build and orchestrate the ETL (Extract, Transform, Load) pipeline.
* Pulls raw data from Azure Data Lake and performs transformation operations (e.g., filtering, cleaning, aggregating).
* Sends the transformed data to Azure Synapse Analytics for further analysis.

1. **Azure Synapse Analytics**

* Receives structured and cleaned data from Azure Data Factory.
* Enables advanced querying, joins, and storage in a SQL-based environment.
* Acts as the core analytical engine that supports trend analysis and diagnostics.

1. **Data Engineer**

* Builds and manages the data pipeline using Azure Data Factory.
* Ensures the data flow is automated, reliable, and scalable.
* Monitors Synapse Analytics outputs to verify pipeline accuracy and performance.

1. **Data Analyst**

* Accesses clean, structured data from Synapse Analytics.
* Performs data exploration and pattern recognition.
* Processes data and creates reports using Power BI for visualization.

1. **Power BI**

* Connects directly to Synapse or imports prepared datasets.
* Enables creation of interactive dashboards and visualizations.
* Translates numerical insights into accessible business intelligence for stakeholders.

1. **Overall Flow**

* Starts from raw data collection ➝ flows through transformation and modeling ➝ ends with rich visualization and business insights.
* Demonstrates a complete data analytics lifecycle using Azure ecosystem tools.

A diagram of a software company

Description automatically generated

**Recommendation**

**1. Capitalize on Seasonal Sales Peaks**

From the monthly sales analysis, it is evident that **June**, **September**, and **December** consistently show **high sales activity** across all years. This indicates strong seasonal demand.

**Recommendations:**

* Launch targeted **promotional campaigns** and **discount offers** in these months to maximize revenue.
* Ensure **inventory planning** is aligned to avoid stockouts during peak months.
* Use **early-month reminders and marketing emails** to capitalize on expected seasonal surges.

**2. Address Q1 Sales Slump**

The months of **January to March** show relatively **lower sales volume** each year. This trend suggests post-holiday fatigue and potential disengagement.

**Recommendations:**

* Introduce **“New Year Kickoff” promotions** to drive engagement in Q1.
* Offer **bundled products**, loyalty rewards, or new product launches to revive customer interest.
* Analyze customer churn in this period and proactively target high-value customers with exclusive offers.

**3. Leverage Top-Performing Regions**

The regional analysis showed that **West and East regions** consistently lead in total sales. However, **Central and South regions underperform** by comparison.

**Recommendations:**

* **Increase advertising spend** and campaign frequency in Central and South to raise brand awareness.
* Conduct **regional preference analysis** to tailor product bundles and pricing more effectively.
* Investigate logistics or delivery challenges that may be affecting customer experience in underperforming areas.

**4. Prioritize High-Value Segments**

The **Consumer segment** remains the strongest contributor to revenue across all years, while **Home Office** lags behind.

**Recommendations:**

* Continue strengthening value propositions and product availability for the **Consumer segment**.
* Explore **cross-sell or upsell opportunities** in Corporate and Home Office segments—e.g., offering B2B bundles or enterprise-level support.
* Consider **segment-specific dashboards** to track behavior and target promotions more precisely.

**5. Improve Data Consistency and Trend Tracking**

Some inconsistencies in sales trends across specific months (e.g., April or August) indicate **possible gaps in marketing alignment, supply chain issues, or irregular campaign performance**.

**Recommendations:**

* Introduce a **seasonal forecasting model** to guide planning for each product category and region.
* Use **moving averages and trend indicators** in dashboards to smooth volatility and detect underlying patterns.
* Establish a **monthly analytics review loop** to catch sudden dips or surges early and adjust strategy.

**6. Expand Visual Insight Depth**

The current set of visualizations offers strong insights on macro trends, but user engagement can benefit from more drill-down and cross-filtering options.

**Recommendations:**

* Add **interactive slicers** by region, segment, and category to enable user-driven exploration.
* Use **forecasting visuals** or trendlines in Power BI to project future movement based on current data.
* Enable **tooltips with KPIs** (profit margin, shipping delay, discount level) to give context at a glance.

**Conclusion**

* This project has been able to clearly illustrate how a modern data analytics pipeline, implemented on Microsoft Azure services, can be used to turn raw business data into usable intelligence. With the help of Azure Data Lake, Azure Data Factory, Azure Synapse Analytics, and Power BI, the sales data of the Superstore dataset has been cleaned, processed, analyzed, and visualised in a unified, scalable environment.
* There are several key insights that the analysis revealed. The seasonal patterns were also easily distinguished with the highest sales recorded in June, September and December of all the years that were studied. These months are the precious sources of revenues which may be used by means of planned marketing and logistics organization. On the other hand, the quarter Q1 (January to March) was constantly performing poorly which implied that the tactics of engagement should be new in order to equalize the sales throughout the year.
* Regional analysis revealed that West and East regions were performing very well, whereas Central and South were still behind, so it would be necessary to have region-specific campaigns, as well as to optimize operations in these areas. Moreover, it was validated that the Consumer segment is the most significant part of the revenue and it can serve as a solid basis of future plans, and the Home Office and Corporate segments can be developed with specific offers and personalization.
* Technically, another good thing about this project is the importance of constructing a cloud-based data pipeline. Azure Synapse facilitated the use of powerful querying and diagnostics and Power BI provided highly impressive visuals that helped identify trends and patterns easily. Interactive dashboards, custom visuals, and measures including moving averages and trendlines delivered a dynamic reporting experience to the stakeholders. Last but not least, the list of recommendations created, including campaign timing and regional focus, dashboard design and predictive analytics, provides the guide to making the analytics of the organization descriptive-to-strategic. When adopted, these measures have the potential to enhance profitability as well as enhance customer responsiveness and operational flexibility.
* Finally, it is possible to note that the report can be used as both a technical demonstration and a business justification of using Azure-powered analytics to make data-informed decisions. It features the value of bringing together technology, data, and strategy to discover valuable insights and provide a quantifiable business effect.