Fundamentals of Data Analysis

Group 8, Class: K20414C

Year: 2021-2022

Lecturer: Nguyen Van Ho (Ph.D.)

University of Economics and Laws

Group Name: The Analysts

List of Member and Task Assessment

No	Name	Student Code	Initial Role
1	Huynh Chi Dung	K204140638	Leader – Thesis composer – Data Analytic Process, presenter
2	Nguyen Ngoc Binh	K204141907	Theoretical Basis validator
3	Pham Duy Tung	K214140961	Theoretical Basis composer, Thesis composer – Overview
4	Tran Nhat Nguyen	K204141923	Theoretical Basis composer, presenter
5	Tran Thu Ha	K204141913	Theoretical Basis composer – BI Tools, Project secretary

Task Assessment

Value Member	Attendance (20%)	Content Requirement (50%)	Progress guarantee (30%)	Total (%)
Huynh Chi Dung	20	50	30	100
Nguyen Ngoc Binh	20	50	30	100
Pham Duy Tung	20	50	30	100
Tran Nhat Nguyen	20	50	30	100
Tran Thu Ha	20	50	30	100

Abstract

Data-driven is an emerging trend that brings the need for more effective data analysis and visualization to get new knowledge and to leverage the benefits of advanced analytics of the volume of data they collect without IT knowledge. Analyzing and visualizing large volumes of data in business often suffers from performance in traditional systems with traditional tools. For understanding the data through visualization, we have tried various approaches, but this described in the paper with Power BI was the most efficient. This paper aims to provide a use case of effective implementation of Power BI tools in analyzing business data, more specifically in the Retail sample database of Microsoft, using the methodology of Big Data analytics and the features of Power BI tool.

Keywords: Data Analysis, Power BI, Descriptive Analytics

Fundamentals of Data Analysis

Contents

List of Abbreviations

Abbreviation	Explanation
IT	the science and activity of using computers and other electronic equipment to store and send information, *Cambridge Dictionary*
BI	Business intelligence (BI) refers to the procedural and technical infrastructure that collects, stores, and analyzes the data produced by a company's activities, *Investopedia*.
KPI	Key performance indicators (KPIs) refer to a set of quantifiable measurements used to gauge a company's overall long-term performance, *Investopedia.*
AI	Artificial intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving, *Investopedia.*

Figures title:

Tables

Table 1

List of tables and fields in Retail Analysis Dataset

Table	Field	Field type	Data type
District	BusinessUnitID	Column	Whole number
District	District	Column	Text
District	DistrictID	Column	Whole number
District	DM	Column	Text / Image
District	DM_Pic	Column	Text
District	DM_Pic_fl	Column	Text
District	BusinessUnitID	Column	Whole number
District	DMImage	Column	Binary / Image
Item	ItemID	Column	Whole number
Item	Segment	Column	Text
Item	Category	Column	Text
Item	Buyer	Column	Text
Item	FamilyName	Column	Text
Sales	MonthID	Column	Whole number
Sales	ItemID	Column	Whole number
Sales	LocationID	Column	Whole number
Sales	Sum_GrossMarginAmount	Calculated Column	Decimal number
Sales	Sum_Regular_Sales_Dollars	Calculated Column	Decimal number
Sales	Sum_Markdown_Sales_Dollars	Calcualted Column	Decimal number
Sales	ScenarioID	Column	Whole number
Sales	ReportingPeriodID	Column	Whole number
Sales	Sum_Regular_Sales_Unit	Column	Whole number
Sales	Sum_Markdown_Sales_Units	Column	Decimal number
Store	LocationID		Whole number/Image
Store	CityName		Text
Store	Territory		Text
Store	PostalCode		Text
Store	OpenDate		Date/Time
Store	SellingAreaSize		Whole number
Store	DistrictName		Text
Store	Name		Text
Store	StoreNumberName		Text
Store	StoreNumber	Calculated Column	Whole number
Store	City	Calculated Column	Text
Store	Chain		Text
Store	DM		Text
Store	DistrictID		Whole number
Store	OpenYear	Calculated Column	Whole number
Store	Store Type	Calculated Column	Text

Table	Field	Field type	Data type
Store	Open Month No		Whole number
Store	Open Month	Calculated Column	Text
Time	ReportingPeriodID		Whole number
Time	Period	Calculated Column	Whole number
Time	FiscalMonth		Text
Time	Month		Date/Time
Time	FiscalYear	Calculated Column	Whole number

Table 2

Primary Keys and Relationships among Tables

Primary Key	From Table	To Table	Relationship Type
DistrictID	District	Store	One-to-many
LocationID	Store	Sales	One-to-many
ReportingPeriodID	Time	Sales	One-to-many
ItemID	Item	Sales	One-to-many

Table 3

Measures used in Dataset

Table	Measure	Formula	Outcome
Sales	Average Unit Price	= [Avg $/Unit TY]	Calculate the Average Unit Price on Total Unit sold of This year
Sales	Average Unit Price Last Year	= [Avg $/Unit LY]	Calculate the Average Unit Price on Total Unit sold of Last year
Sales	Avg $/Unit LY	= IF([Total Units Last Year]<>0, [TotalSalesLY]/[Total Units Last Year], BLANK())	Calculate the Average Sales on Total Unit sold Last Year. Leave blank if there is no unit sold.
Sales	Avg $/Unit TY	= IF([Total Units This Year]<>0, [TotalSalesTY]/[Total Units This Year], BLANK())	Calculate the Average Sales on Total Unit sold This Year. Leave blank if there is no unit sold.
Sales	Gross Margin Last Year	= CALCULATE(SUM([Sum_GrossMarginAmount]), Sales[ScenarioID]=2)	Calculate the Gross Margin Last Year by Aggregating the gross margin amount with ScenarioID =2 for Last Year.
Sales	Gross Margin	= [Gross Margin Last Year]/[TotalSalesLY]	Calculate Gross Margin Last Year % by using Gross

Table	Measure	Formula	Outcome
	Last Year %		Margin Last Year divide by Total Sales Last Year, Data type is percentage.
Sales	Gross Margin This Year	= CALCULATE(SUM([Sum_GrossMarginAmount]), Sales[ScenarioID]=1)	Calculate the Gross Margin This Year by Aggregating the gross margin amount with ScenarioID =1 for This Year.
Sales	Gross Margin This Year %	= [Gross Margin This Year]/[TotalSalesTY]	Calculate Gross Margin This Year % by using Gross Margin This Year divide by Total Sales This Year, Data type is percentage.
Sales	Last Year Sales	= [TotalSalesLY]	Sum of Last Year Sales
Sales	Markdown_Sales_Dollars	= SUM([Sum_Markdown_Sales_Dollars])	
Sales	Markdown_Sales_Units	= SUM([Sum_Markdown_Sales_Units])	
Sales	Regular_Sales_Dollars	= SUM([Sum_Regular_Sales_Dollars])	
Sales	Regular_Sales_Units	= SUM([Sum_Regular_Sales_Units])	
Sales	Sales Per Sq Ft	= ([TotalSalesTY]/(DISTINCTCOUNT([MonthID])*SUM(Store[SellingAreaSize])))*12	Calculate Sales per Square Foot by of Each store over the year using the sum sales per month.
Sales	Store Count	= DISTINCTCOUNT([LocationID])	Indicate the number of stores using Location ID
Sales	Total Sales Var	= [TotalSalesTY]-[TotalSalesLY]	Calculate the sales variance by using this year sales subtract last year sales.
Sales	Total Sales Var %	= IF([TotalSalesLY]<>0, [Total Sales Var]/[TotalSalesLY], BLANK())	Calculate the % of Variance, leave blank if sales last year =0
Sales	Total Sales Variance	= [Total Sales Var]	
Sales	Total Sales Variance %	= [Total Sales Var %]	
Sales	Total Units Last Year	= CALCULATE([TotalUnits], Sales[ScenarioID]=2)	Calculate the total units sold with SalesScenarioID =2 for Last Year.

Table	Measure	Formula	Outcome
Sales	Total Units This Year	= CALCULATE([TotalUnits], Sales[ScenarioID]=1)	Calculate the total units sold with SalesScenarioID = 1 for This Year.
Sales	TotalSales	= [Regular_Sales_Dollars]+[Markdown_Sales_Dollars]	Calculate the Total Sales by total the Regular Sales Dollars and Markdown Sales Dollars
Sales	TotalSales LY	= CALCULATE([TotalSales], Sales[ScenarioID]=2)	Calculate the Total Sales with SalesScenarioID = 2 for Last Year
Sales	TotalUnits	= [Regular_Sales_Units]+[Markdown_Sales_Units]	Calculate the Total Units by total the Regular Sales Units and Markdown Sales Units
Store	Average Selling Area Size	= AVERAGE([SellingAreaSize])	
Store	Count of OpenDate	= COUNTA('Store'[OpenDate])	Count all the value <> blank of the open dates.
Store	New Stores	= CALCULATE(COUNTA([Store Type]), FILTER(ALL(Store), [Store Type]="New Store"))	Count the dataset filtered with New Store
Store	New Stores Target	= 14	Set target for number of new stores of each chain distinctly counted.
Store	Open Store Count	= COUNTA([OpenDate])	

Chapter 1: An Overview of the Thesis

The reason for choosing the topic

The justification for choosing the topic is the desire to apply the learned data analysis technique to a set of available sample data, thereby extracting helpful information and insights, research, analyzing the current problems of the business, as well as proposing grounded solutions. Supporting and optimizing the decision-making of the business in the future.

Topic goal

This study was conducted with the aim of generalizing and positioning data analytics science nowadays, and concurrently illustrating the entire process of performing data analysis for a particular project through application of descriptive analytic methods.

Subject and research scope of the project

Subject of the paper is Descriptive Analytics in Data Analysis, the project uses qualitative method with case study.

Tools used

Microsoft Power BI: Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights, which will be the primary tool to be implemented in this project.

Method of analyzing and selecting case study data:

- - Step 1: Plan - choose a method
- - Step 2: Research design
- - Step 3: Prepare before conducting data collection
- - Step 4: Collect data

- - Step 5: Analyze the data using the designated analytic method
- - Step 6: Share and discuss analysis results

Each of the above steps can be cross related throughout the entire process, not just the next step. Research design needs to be adjusted when errors occur during implementation, so the flexibility of the researcher is essential.

Research implications

This study was conducted with the aim of providing a comprehensive view of the science of data analysis, as well as the application of descriptive analysis to an organization's data set. Moreover, the article also points out the limitations and recommendations of the research results/research methods, so that the reader can have a harvest from the most objective perspective.

Structure of report

The problem is addressed in two related parts: The problem statement is contained in Chapter 1, and a review of the related research, theory, and professional literature are described in Chapter 2. The methods used for investigating the problem are usually included in Chapter 3. The outcomes are traditionally presented in two chapters: Chapter 4 is devoted to reporting the results, and Chapter 5 presents the conclusions and implications drawn from the results.

- In Chapter 1: Overview of thesis. A compelling case is made regarding the problem under investigation, the purpose of the study, and the research questions to be investigated.
- Chapter 2: Theoretical basis. This chapter serves as the foundation on which the study is built and as a basis for discussing results and interpretations.

- Chapter 3: Analysis of user requirements and data description. In this chapter, user requirements will be identified using the method of analyzing and selecting case study data. Describing, understanding the source database, and selecting data relevant to user requirements.
- Chapter 4: Data analysis and results. In this chapter, a factual reporting of the study results is presented. Findings are organized around the research questions. The readers will learn what techniques and tools the research plan to use to analyses and summarize the data. Depending upon the user requirements, the data sources and data analyses will be used to answer each research question.
- Chapter 5: Conclusion. In this final chapter, the implications of the study findings are discussed. Findings are integrated with the theory employed in the first chapter and the body of knowledge presented in the second chapter. The chapter ends with a cogent conclusion summarizing the importance of the study findings.

Chapter 2: Theoretical Basis

2.1. Overview of Data Analytics. Graduate Programs Staff, *How Do I Start a Data Project: Understanding a Project Lifecycle*, August 2020

2.1.1. The benefits of Data Analytics in the business.

What is Data Analytics?

- Data analytics is a process of examining, cleaning, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision - making.
- Any sort of information can be subjected to data analytics techniques to gain insight that can be used to improve things. Data analytics approaches can uncover trends and

indicators that would otherwise be lost in the shuffle of data. This data can then be utilized to optimize procedures and increase the overall efficiency of a business or system.

The benefits of Data Analytics in the business.

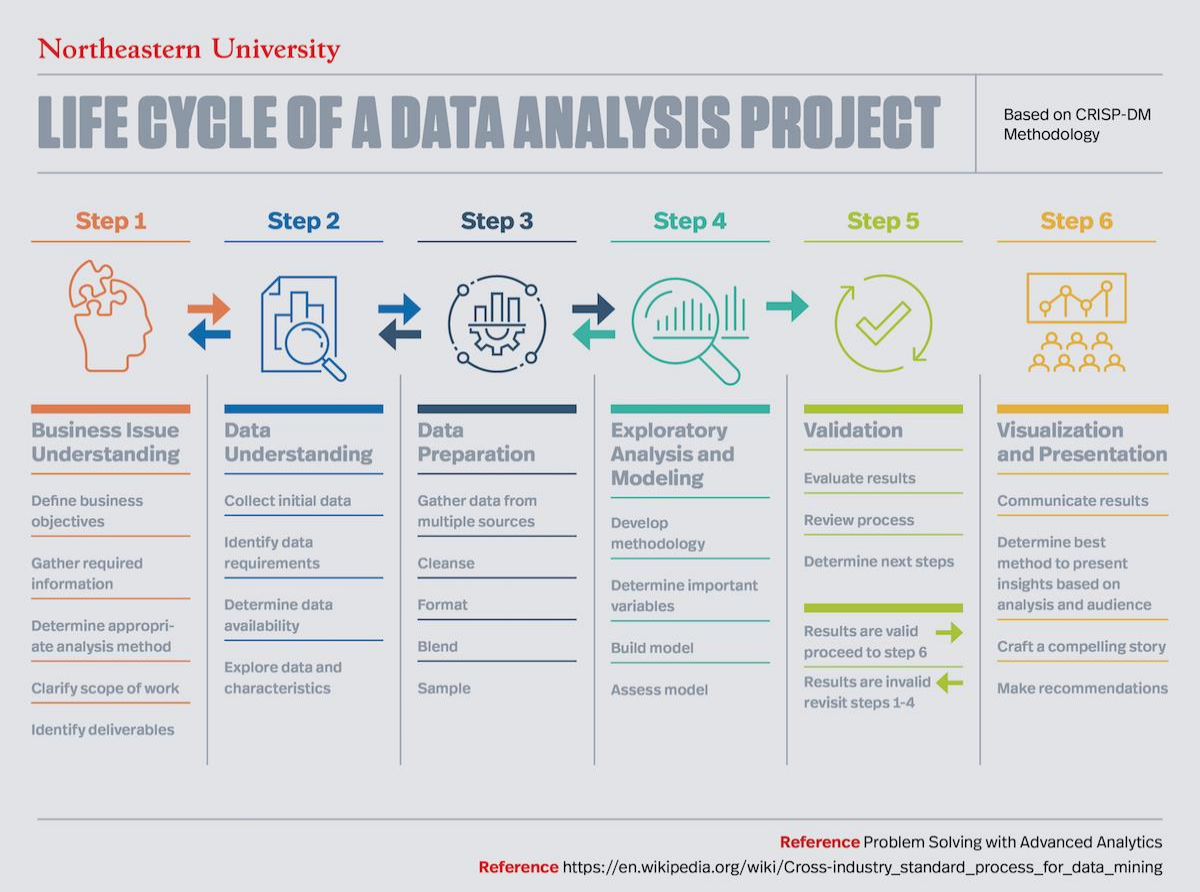
- In today's business arena, data analytics plays a role in helping to make scientific decisions and help businesses operate more efficiently, optimize their performances.
- Data analytics is a powerful decision - making tool that allows business leaders to have all the information they need to move their organization in the right direction.
- Implementing data analytics into the business model can help reduce costs by identifying more efficient ways of doing business and by storing large amounts of data.
- Customers' data is collected through a variety of channels, including physical retail, e-commerce, and social networking. Businesses can get insights into consumer behavior and provide a more personalized experience by using data analytics to create comprehensive customer profiles from this data, analyze customer trends and satisfaction, leading to new - and better - products and services.
- Data analytics can assist a company in identifying dangers and taking preventative steps. A retail chain, for example, could use a propensity model - a statistical model that predicts future behaviors or events - to figure out which outlets are most vulnerable to theft. The company might then use this information to decide the level of security required at each location, as well as if it should divest from any of them.

- Furthermore, using data analytics can be one of the most effective ways to strengthen your organization's cybersecurity and prevent serious threats from causing either financial or reputational damage.

2.1.2. The process of doing data analytics project in business.

Figure 1:

Life Cycle of a Data Analysis Project



Reference: Problem Solving with Advanced Analytics, Northeastern

Step 1: Understand the Business Issues

When presented with a data project, you will be given a brief outline of the expectations. From that outline, you should be able to figure out what the company's

main goals are. You should examine the overall scope of the work, business objectives, information the stakeholders are seeking, the type of analysis they want you to use, and the deliverables (the outputs of the project) they want.

Step 2: Understand Your Data Set

There are a variety of tools you can use to organize your data. When presented with a small dataset, you can use Excel, but for heftier jobs, you will likely want to use more rigid tools to explore and prepare your data. R, Python, Alteryx, Tableau Prep or Tableau Desktop, Power BI can help prepare your data for its cleaning.

Within these programs, you should identify key variables to help categorize the data. When going through the data sets, look for errors in the data. These can be anything from omitted data, data that does not logically make sense, duplicate data, or even spelling errors. These missing variables need to be amended so you can properly clean your data.

Step 3: Prepare the Data

Once you have organized and identified all the variables in your dataset, you can begin cleaning. In this step, you will input missing variables, create new broad categories to help categorize data that does not have a proper place, and remove any duplicates in your data. Imputing average data scores for categories where there are missing values will help the data be processed more efficiently without skewing it.

Step 4: Perform Exploratory Analysis and Modeling

In this step, you will begin building models to test your data and seek out answers to the objectives given. Using different statistical modeling methods, you can determine

which is the best for your data. Common models include linear regressions, decision trees, and random forest modeling, among others.

Step 5: Validation

Once you have crafted your models, you will need to assess the data and determine if you have the correct information for your deliverable. Did the models work properly? Does the data need more cleaning? Did you find the outcome the client was looking to answer? If not, you may need to go over the previous steps again.

Step 6: Visualize and Presentation

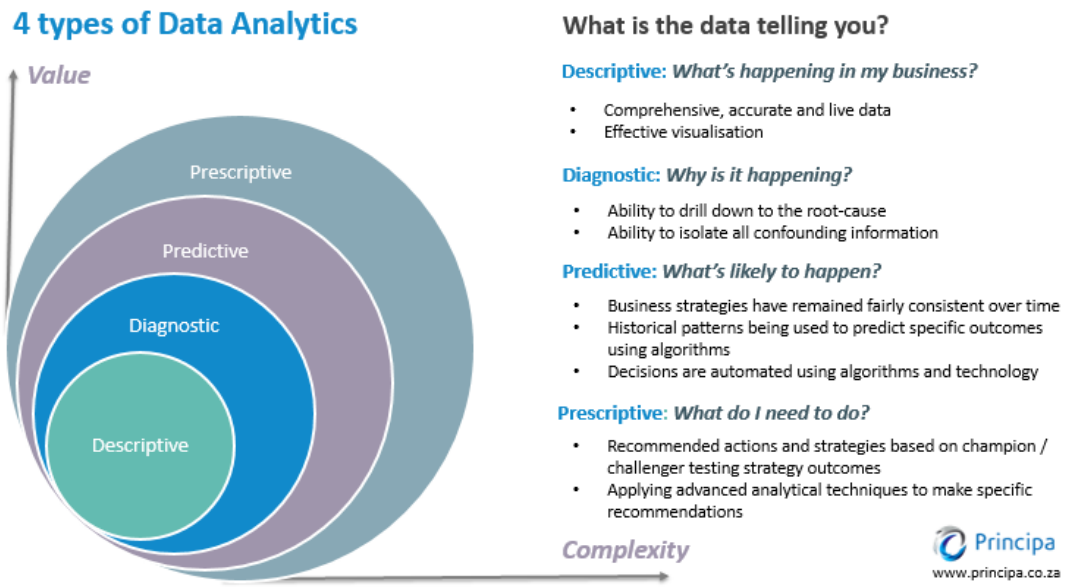
Once you have all your deliverables met, you can begin your data visualization. In many cases, data visualization will be crucial in communicating your findings to the client. Not all clients are data-savvy, and interactive visualization tools like Tableau, Power BI are tremendously useful in illustrating your conclusions to clients. Being able to tell a story with your data is essential. Telling a story will help explain to the client the value of your findings.

2.2. Theory and methods in data analysis.

2.2.1. Type of data analysis.

Figure 2:

Four Types of Data Analytics



Reference: www.principa.co.za

Modern analytics tend to fall in four distinct categories:

- Descriptive Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

The following will introduce each type and give examples of how they are utilized in business.

Descriptive Analysis.

Definition: Descriptive analytics is the interpretation of historical data to better understand changes that have occurred in a business. Descriptive analytics describes the use of a range of historic data to draw comparisons. Most reported financial metrics are a product of descriptive analytics. Descriptive analysis answers the "what happened" by summarizing past data, usually in the form of dashboards.

Roles and Impacts: Descriptive analysis is at the foundation of all data insight. It is the simplest and most common use of data in business today.

The biggest use of descriptive analysis in business is to track Key Performance Indicators (KPIs). KPIs describe how a business is performing based on chosen benchmarks.

Example of descriptive analysis:

- KPI dashboards
- Monthly revenue reports
- Sales lead overview.

This type of analysis will be conducted in this project.

Diagnostic analysis.

Definition: Diagnostic analytics is a form of data analytics that builds on descriptive analytics to help you understand why something happened in the past. Often, diagnostic analysis is referred to as root cause analysis. It involves processes such as data discovery, data mining, and drill down and drill through.

Roles and Impacts: Diagnostic Analysis takes the insights found from descriptive analytics and drills down to find the causes of those outcomes. Organizations use this type of analytics as it creates more connections between data and identifies patterns of behavior.

A critical aspect of Diagnostic Analysis is creating detailed information. When new problems arise, it is possible you have already collected certain data pertaining to the issue. By already having the data at your disposal, it ends having to repeat work and makes all problems interconnected.

Example of diagnostic analysis:

- A freight company investigating the cause of slow shipments in a certain region
- A SaaS company drilling down to determine which marketing activities increased trials.

Predictive Analysis.

Definition: The term predictive analytics refers to the use of statistics and modeling techniques to make predictions about the outcomes and performance. Predictive analytics looks at current and historical data patterns to determine if those patterns are likely to emerge again. Predictive Analysis attempts to answer the question "what is likely to happen". This type of analytics utilizes previous data to make predictions about the future situation.

Roles and Impacts: This type of analysis is another step up from the descriptive and diagnostic analyses. Predictive analysis uses the data we have summarized to make logical predictions of the outcomes of events. This analysis relies on statistical modeling, which requires added technology and manpower to forecast. It is also essential to understand that forecasting is only an estimate; the accuracy of predictions relies on the quality and detailed data.

Example of predictive analysis:

- Risk Assessment
- Sales Forecasting
- Using customer segmentation to determine which leads have the best chance of converting

- Predictive analytics in customer success teams.

Prescriptive Analysis.

Definition: Prescriptive Analytics factors information about possible situations or scenarios, available resources, past performance, and current performance, and suggests a course of action or strategy. It can be used to make decisions on any time horizon, from immediate to long-term. Prescriptive Analysis is the frontier of data analysis, combining the insight from all previous analyses to determine the course of action to take in a current problem or decision.

Roles and Impacts: The opposite of prescriptive analytics is descriptive analytics, which examines decisions and outcomes after the fact.

Example of prescriptive analytics: Artificial Intelligence (AI). AI systems consume a large amount of data to continuously learn and use this information to make informed decisions. Well-designed AI systems can communicate these decisions and even putting those decisions into action. Business processes can be performed and optimized daily without a human doing anything with artificial intelligence. Currently, the biggest data-driven companies (Apple, Facebook, Netflix, etc.) are utilizing prescriptive analytics and AI to improve decision-making.

2.2.2. An Overview of Descriptive Analysis.

Descriptive Analysis is the type of analysis of data that helps describe, show, or summarize data points in a constructive way such that patterns might emerge that fulfill every condition of the data.

Techniques for Descriptive Analysis: Data aggregation and data mining are two techniques used in descriptive analysis to churn out historical data. In Data aggregation, data is first collected and then sorted to make the datasets more manageable.

Types of Descriptive Analysis:

Descriptive analysis can be categorized into four types which are measures of frequency, central tendency, dispersion or variation, and position. These methods are optimal for a single variable at a time.

- *Measures of Frequency:* The frequency of a certain event or response is likely to occur. This is the prime purpose of measures of frequency to make something like a count or percent.
- *Measures of Central Tendency:* Finding out the Central (or average) Tendency or response. Central tendency is measured with the use of three averages — mean, median, and mode.
- *Measures of Dispersion:* How data is divided across a range.
- *Measures of Position:* Identifying the position of a single value or its response in relation to others. Measures like percentiles and quartiles become very useful in this area of expertise.

Chapter 3: Analysis of User Requirements and Data Description

3.1. Identify and analyze user requirements

- Yearly Sales Overview: based on the given database, Provide an overview of Sales on chain wise. Compare to last year sales to asset chain performance.
- Manager's Performance: Build a sales analysis filtered by district managers segmented on categories wise to assess his performance over the year.

- Top 10 Categories: Call out 10 top performed categories to analyze their impacts on revenue and production.
- New Store Analysis: Provide an analysis on new stores opening this year of both two chains. Overview their performance and scales of expansion.
- Variations of Sales per District: A study on variances of sales per fiscal months.

3.2. Overview of the source database

The team's case study is based on the Retail Analysis dataset of obviEnce® which has been anonymized, published on Microsoft with a package built to be compatible with the Power BI Desktop application. This industry sample dashboard and underlying report analyzes retail sales data of items sold across multiple stores and districts. The metrics compare this year's performance to last year's in sales, units, gross margin, and variance, as well as new store analysis.

Figure 3:

ObviEnce's logo

Reference: ObviEnce.com

ObviEnce is an ISV and an Intellectual Property (IP) Incubator focused on Microsoft Business Intelligence. ObviEnce works closely with Microsoft to develop best practices and thought leadership for jump-starting and deploying Microsoft Business Intelligence solutions.

3.2.1. Describe the source data

This file and associated data is property of obviEnce LLC and has been shared solely for the purpose of demonstrating Power BI functionality with industry sample data.

The dataset is divided into 5 main tables[1]

- District: This table contains information about the districts, affiliated stores, and information and pictures of the county manager.
- Item: The list of items, their categories, segments, and buyers.
- Sales: Data about the organization's revenue
- Store: Detailed data about stores, from locations to chains and store managers. Also note the opening time of the store.
- Time: Data table about the date and time of the report.

3.2.2 Select and present the data to be analyzed for user requirements

- Yearly sales overview on chain wise: by using the consolidated sales data of each store grouped by district data, we can see a picture of the sales over the year.
- Assessing manager’s monthly performance: filter the consolidated sales data of each store by district manager. Data used for this requirement will be District Manager, Stores, Fiscal Year.
- Identify the top 10 sales per year by category: Pulling the consolidated sales data and filter by top 10 categories. Then apply the time series onto the result.

---

[1] Table 1: List of tables and fields in Retail Analysis Dataset

- Viewing new stores development by chain and location: Based on the opening date of each store, summarize the new store counts of each chain on geographical map.
- Viewing total sales variance per district: Calculate the Total Sales Variance %, by Fiscal Month filter by District Manager.

Chapter 4. Data analysis and results

4.1 Introduction to data analysis tools and solutions

What is Power BI?

- Power BI is the collective name for an assortment of cloud-based apps and services that help organizations collate, manage, and analyze data from a variety of sources, through a user-friendly interface.
- Business intelligence tools like Power BI can be used for a multitude of purposes.
- Primarily, Power BI pulls data together and processes it, turning it into intelligible insights, often using visually compelling and easy-to-process charts and graphs. This allows users to generate and share clear and useful snapshots of what's happening in their business.
- Power BI connects to a range of data sources, from basic Excel spreadsheets to databases, and both cloud-based and on-premises apps.
- Power BI is something of an umbrella term and can refer to either a Windows desktop application called Power BI Desktop, an online SaaS (Software as a Service) service called Power BI Service, or mobile Power BI apps available on Windows phones and tablets, as well as for iOS and Android devices.
- Power BI is built on the foundation of Microsoft Excel, and as such, the learning curve from Excel to Power BI is not that steep; anyone who can use Excel can use Power BI, but the latter is far more powerful than its spreadsheet counterpart.

What does Power BI do?

- Microsoft Power BI is used to run reports and surface insights based on a company's data. Power BI can connect to a wide range of data sets, and "tidies

up" the info it is fed so that it can be better digested and understood. The reports and visuals generated from this data can then be shared with other users.

- Power BI helps users see not only what has happened in the past and what is happening in the present, but also what might happen in the future. Power BI is infused with machine learning capabilities, meaning it can spot patterns in data and use those patterns to make informed predictions and run "what if" scenarios. These estimates allow users to generate forecasts and prepare themselves to meet future demand and other key metrics.

Why do people use Power BI?

For businesses who want more reporting power and analytical muscle than that offered by Excel, Power BI is the next level in business intelligence. With Power BI, businesses can assemble, scrutinize, and visualize data from across the company, giving them greater insight into their operations and performance, and allowing them to make more informed decisions based on real data.

The key benefits of using Power BI:

- Businesses can input huge quantities of data into Power BI that many other platforms would struggle to process
- Built-in machine learning features can analyze data and help users spot valuable trends and make educated predictions
- Information can be visualized using powerful templates to allow businesses to better make sense of their data
- Power BI is cloud-based, so users get cutting edge intelligence capabilities and powerful algorithms that are updated regularly

- Powerful personalization capabilities allow users to create dashboards so they can access the data they need quickly
- Alerts can be set up on KPIs to keep users up to date important metrics and measurements.
- Power BI has an intuitive interface that makes it far more user-friendly and easier to navigate then complex spreadsheets
- The platform integrates with other popular business management tools like SharePoint, Office 365, and Dynamics 365, as well as other non-Microsoft products like Spark, Hadoop, Google Analytics, SAP, Salesforce, and MailChimp
- With data security a massive talking point for modern businesses, Power BI ensures data is safe, offering granular controls on accessibility both internally and externally

What components make up Power BI?

- The Power BI product is made up of several apps, each with their own features and uses. These include:
- Power Query: a data connection tool that lets you transform, combine, and enhance data from several sources
- Power Pivot: a data modeling tool for creating data models
- Power View: a data visualization tool that generates interactive charts, graphs, maps, and other visuals
- Power Map: another visualization tool for creating immersive 3D visuals
- Power Q&A: a question-and-answer engine that lets you ask questions about your data in plain language

4.2. Analyze, explore, and visualize data

4.2.1. Data Analyzation

After walking through the list of tables provided in the data, we start to study and establish the relationship between them using primary keys. So, we can connect the tables together to begin the analysis process.[2]

Besides, we also need to transform and model the data to meet the user requirements. We are using the following measurements provided by PowerBI:

- Calculated Columns: Calculated column data is stored in x-Velocity in-memory database. The calculation is made before the model is queried by the user. Calculated columns are good because there is a smaller virtual memory requirement when the user is interacting with the reports. However, they take up more storage in the database.
- Measures: Measures are used to calculated aggregates like Sum or Average. Measures are created at the time of the query, so they are not stored in the database. Supported in DAX Calculation.[3]

4.2.2. Data Exploration

As per the nature of the requirements, we need to calculate the following elements to support for Retail Analysis:

- Sales per Square foot: Sales per square foot is your store's average revenue for every foot of sales space, including non-selling space such as your stock room, fitting room,

---

[2] Table 2: Primary Keys and Relationships among Tables
[3] Table 3: Measures used in Dataset

and receiving areas. Analysts base on this data to calculate the store performance, HR planning, resources, and projects planning.

- Average spending: Analyst calculate this element to plan the capitalization for the business, by developing promotions and marketing.

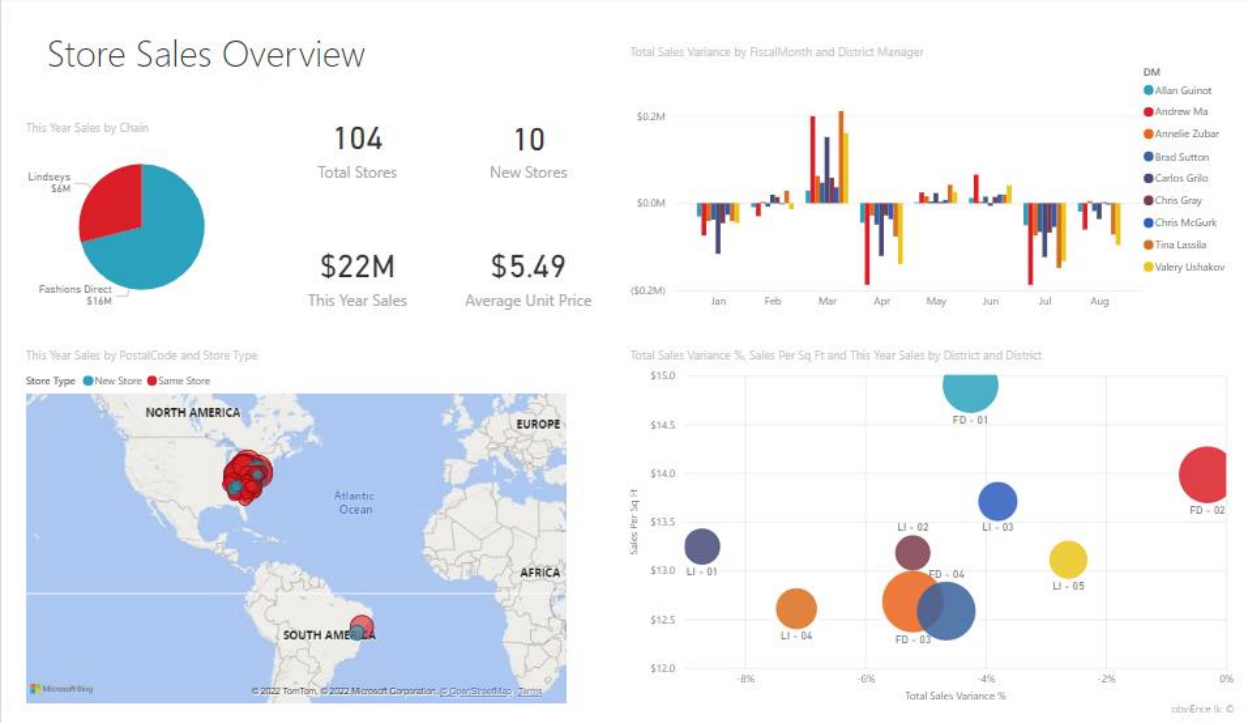
- Variances: refers to a statistical measurement of the spread between numbers in a data set. More specifically, variance measures how far each number in the set is from the mean and thus from every other number in the set. Analysts use this field to assess the performance of the business.

4.2.3. Data Visualization

The dataset is display into the following dashboards to support the descriptive analytics:

Figure 4:

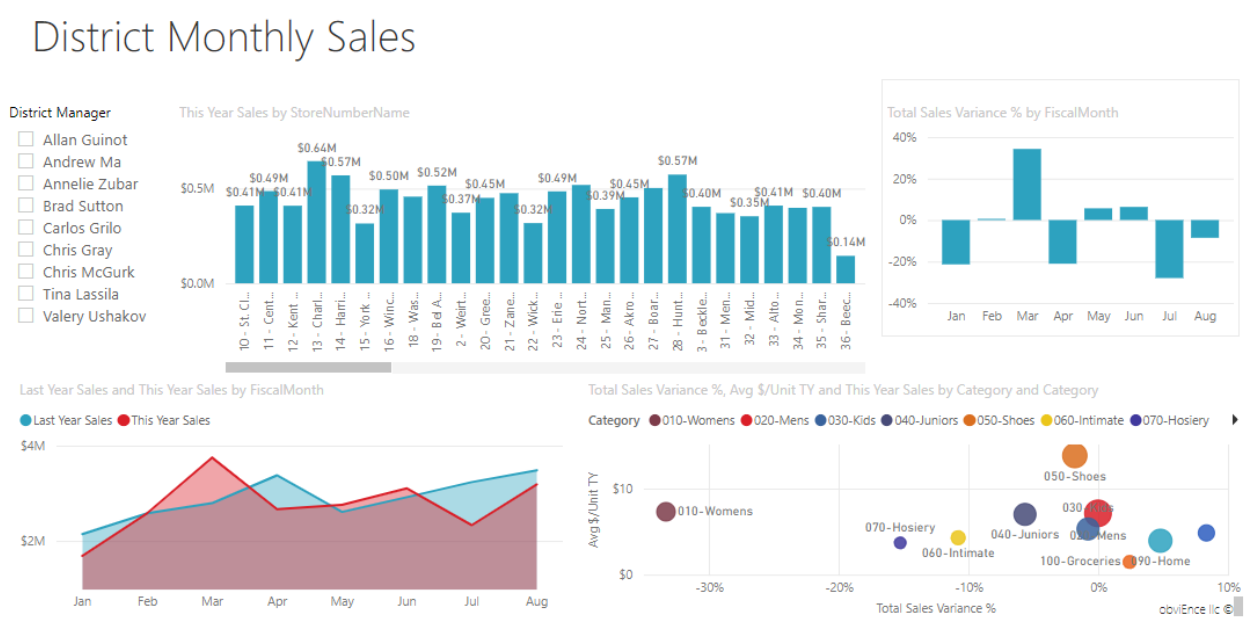
Store Sales Overview Dashboard



This Dashboard includes two main elements for stores revenue by chain and sales variance per district managers. Analyst highlighted the summary by using labels, ratio of chains by using pie chart. Revenue of each store is demonstrated by map view using store locations to support geographical analysis. Sales per Square foot by district is also put into a bubble plot chart.

Figure 5:

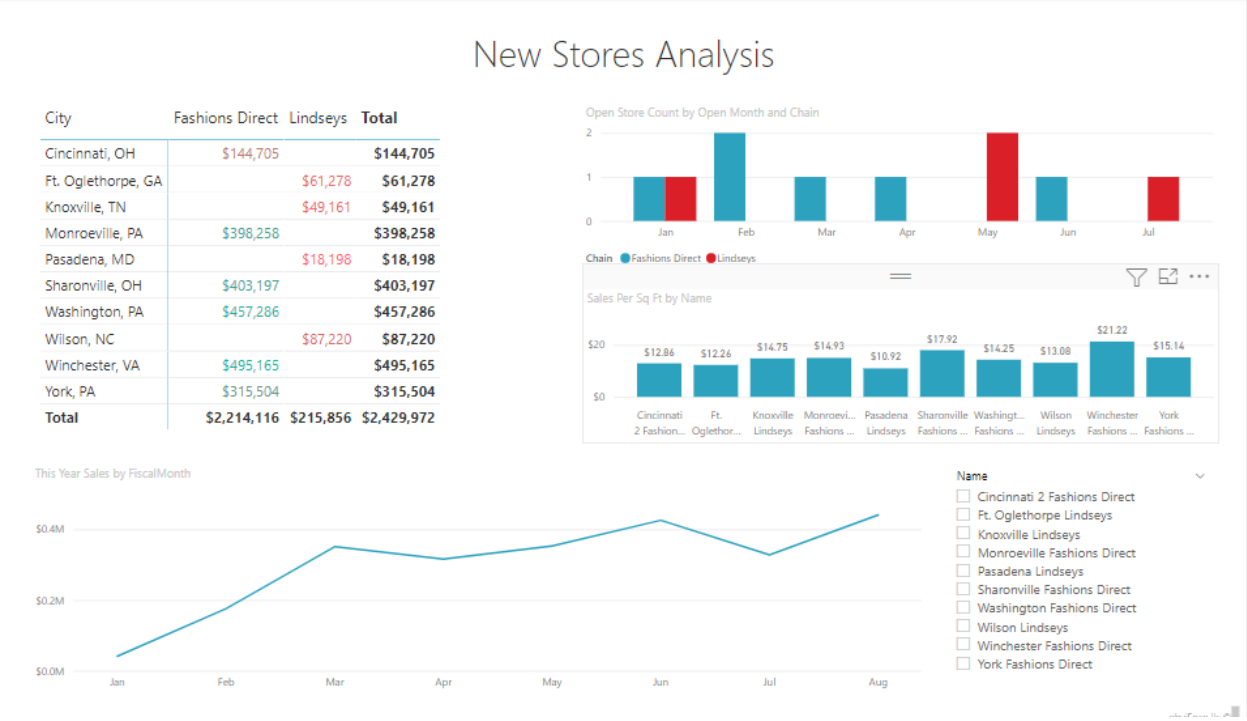
District Monthly Sales Dashboard



District Mangers filed is used as a slicer to analyze sales per category. Sales variance comparing to last year also highlighted. Using the bubble chart to analyze district's manager sales per category for a deep analytic of manager's performance.

Figure 6:

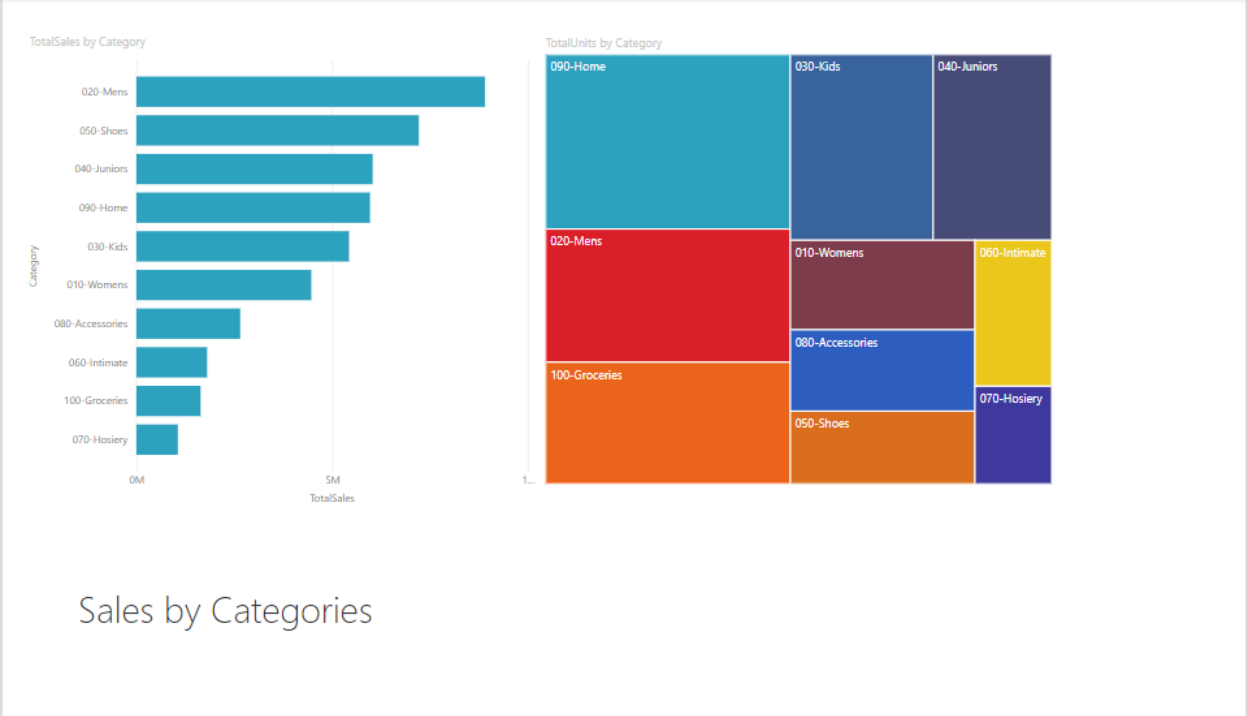
New Store Analysis Dashboard



This Dashboard highlights the sales of new store (opened this year with ScenarioID =1), revenue matrix segmented by chain and detailed by bar charts. Sales of each stores is also listed out for a deeper analytic, supported by a line chart for monthly performance analysis.

Figure 7:

Sales by Categories Dashboard



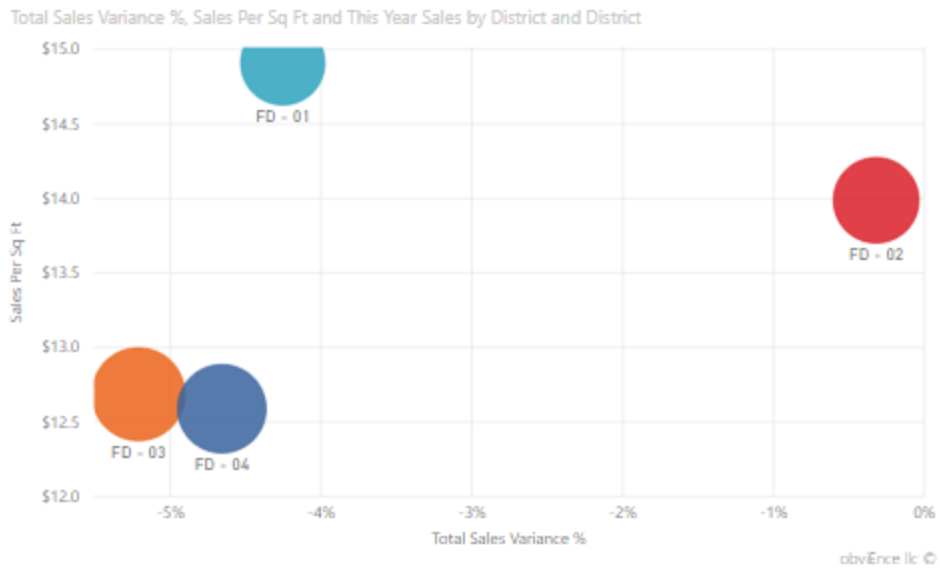
The dashboard calls out top 10 categories sold, by revenue and unit sold to study on the impacts to the overall performance.

4.3 Discuss and evaluate the results, implications from results to support decision making in business

In Sales Overview Dashboard, we can see that Fashion Direct chain takes the most of the pie from the total sales. We start to study this chain deeper:

Figure 8:

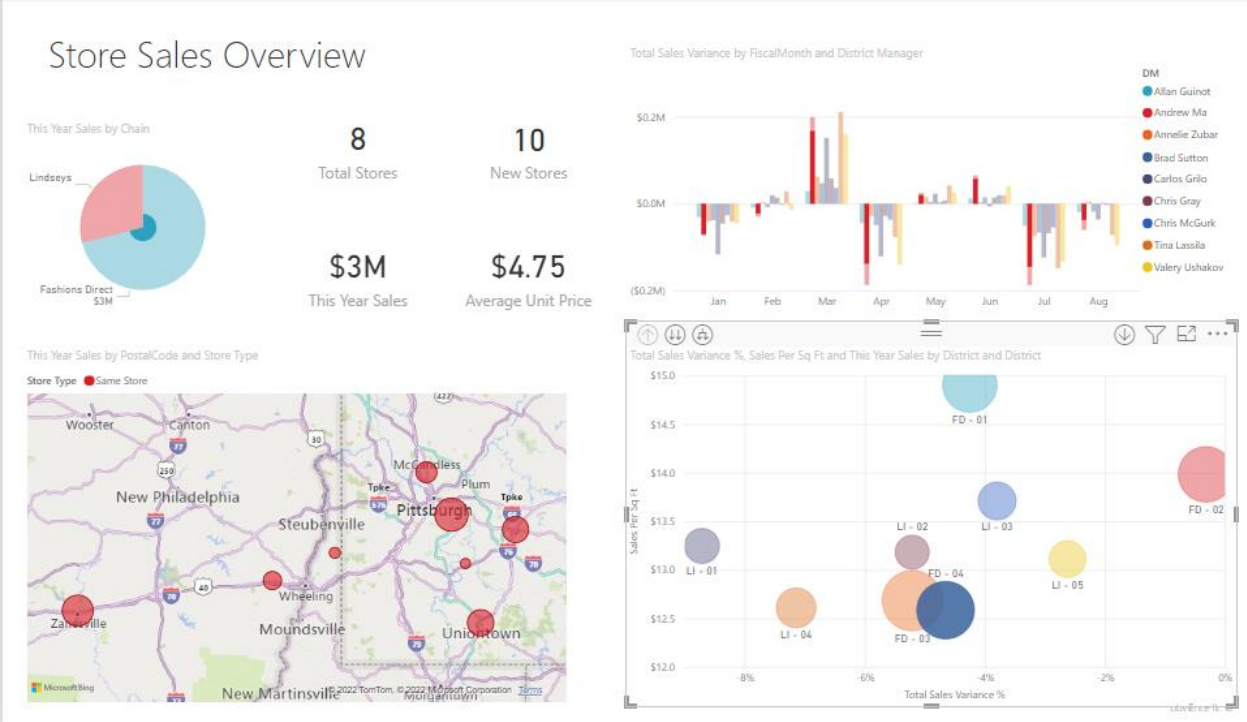
District Sales Overview filtered by Fashion Direct Chain



The FD-01 district has the highest average Sales per Square Foot and FD-02 has the lowest Total Sales Variance compared to last year. FD-03 and FD-04 are worst performers overall.

Figure 9:

District FD-04's Performance

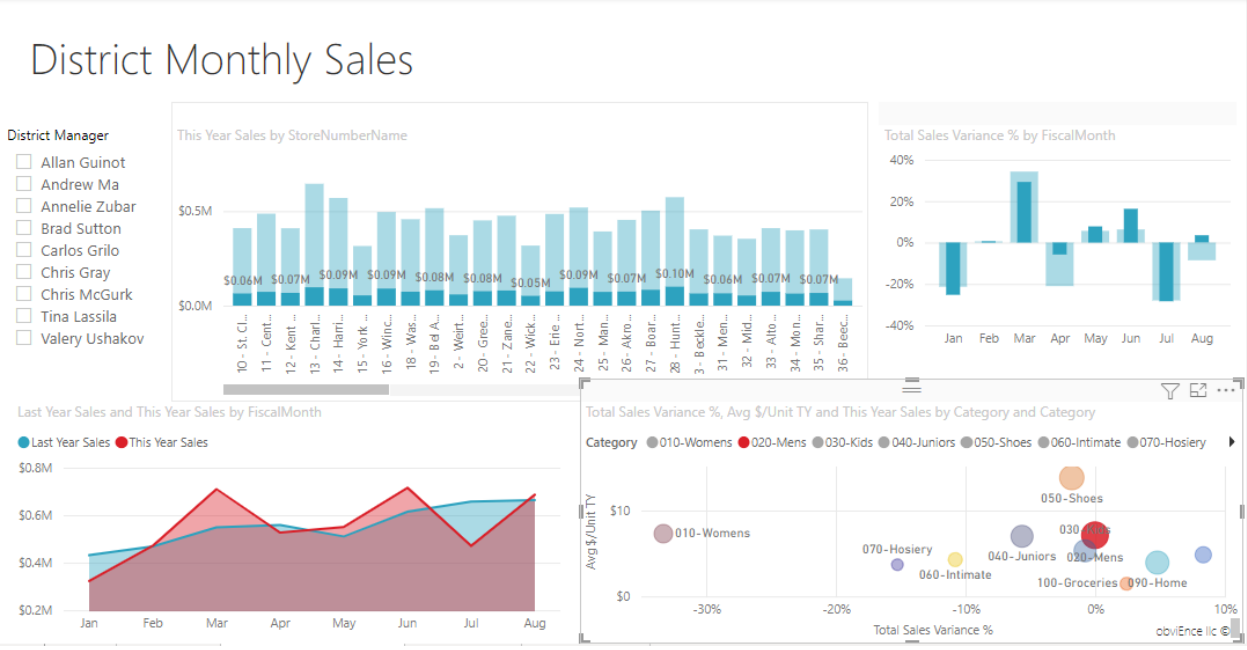


Analyst also concludes that this unit's performance has impacted the most negative variances in the chart. We can base on this result to analyze Andrew Ma (District Manager)'s performance on the next section.

We dive in District Monthly Sales Dashboard, In the Total Sales Variance % by Fiscal Month chart, notice the large variability on variance % compared to last year, with January, April, and July being particularly bad months.

Figure 10:

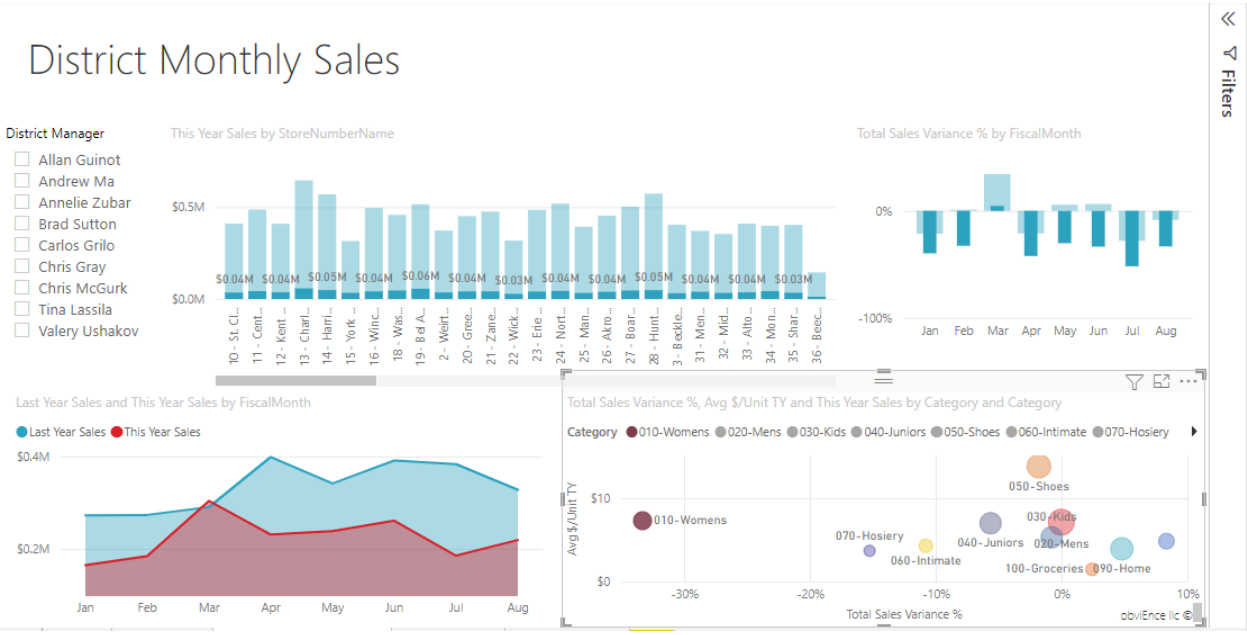
Performance of 020-Mens category



Observe that although the men's category wasn't as severely affected in April as the overall business, January and July were still problematic months.

Figure 11:

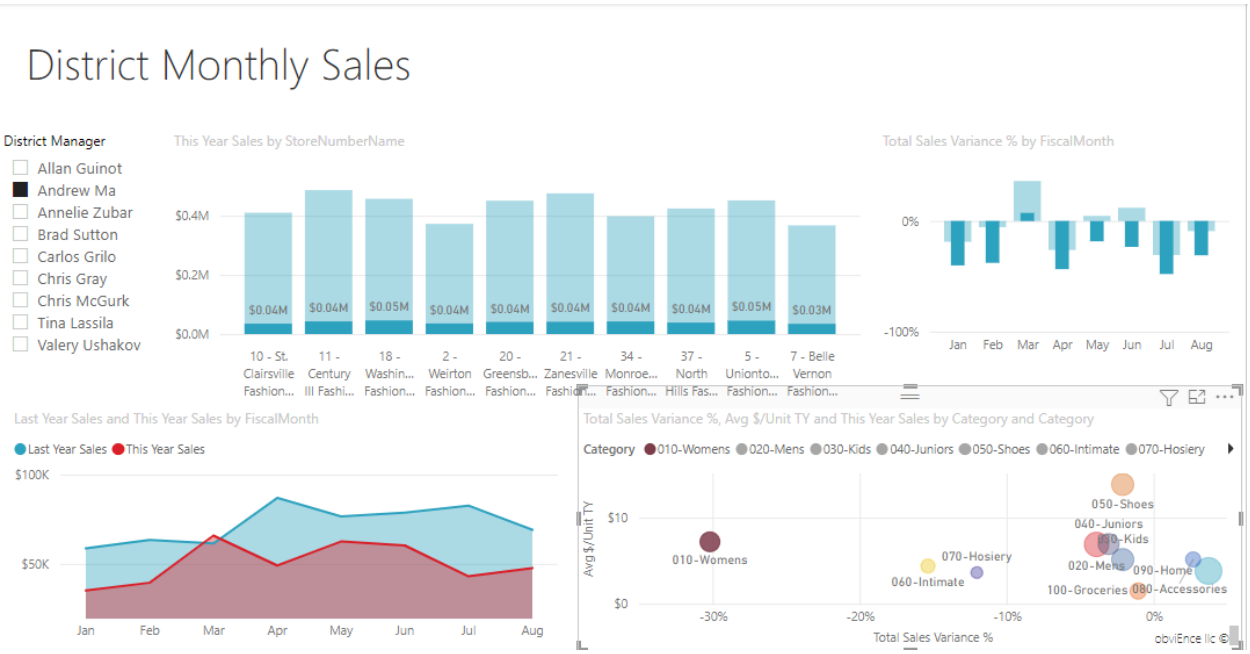
Performance of 010-Womens category



Notice the women's category performed much worse than the overall business across all months, and in almost every month compared to the previous year.

Figure 12:

Andrew Ma's Sales of 010-Womens category



We can see that Andrew did not do well on this category. And this category is not a well performed one compared to the other in both years.

We start to study New Store Development by viewing New Store Analysis Dashboard, As evident from the tile, more Fashions Direct stores than Lindseys stores opened this year. Notice the difference in average sales/square foot across the new stores by observing sales per sq ft by Name chart.

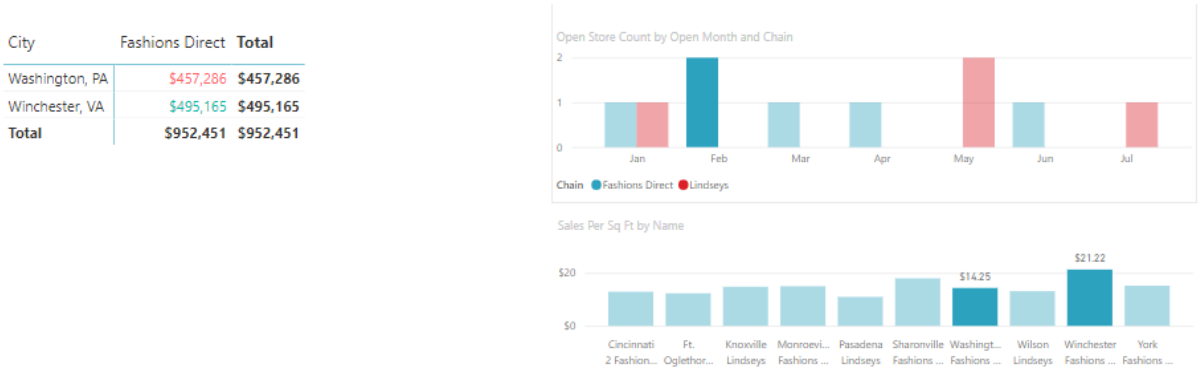
Figure 13:

Sales per Sq Ft by Name chart in New Store Analysis Dashboard

When selecting the Fashions Direct legend item in the Open Store Count by Open Month and Chain top-right chart. Notice, even for the same chain, the best store (Winchester Fashions Direct) significantly outperforms the worst store (Cincinnati 2 Fashions Direct) by $21.22 vs $12.86, respectively.

Figure 14:

Comparison of new stores in Fashion Direct chain



We analyze the Sales by Categories Dashboard. Notice that the 090-Home category takes the most percentage in Unit sold but only stays at the 4th of Sales. Concurrently, 020-Mens category brings the most revenue but only the 2nd in unit consumption.

Figure 15:

090-Home category's performance

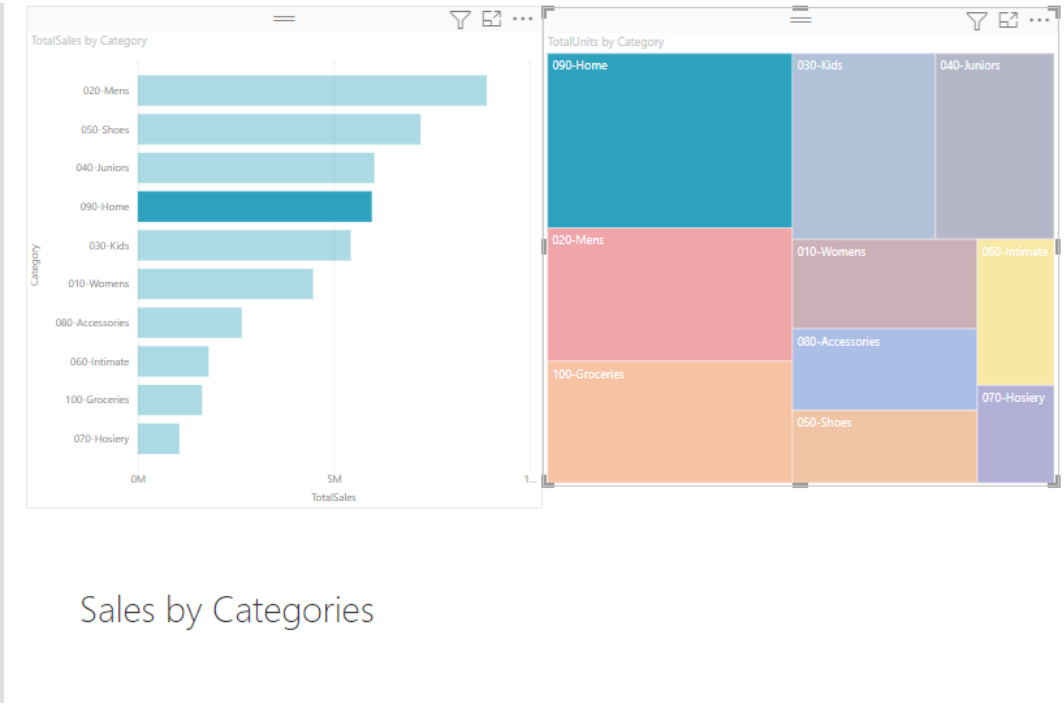


Figure 16:

020-Mens category's performance

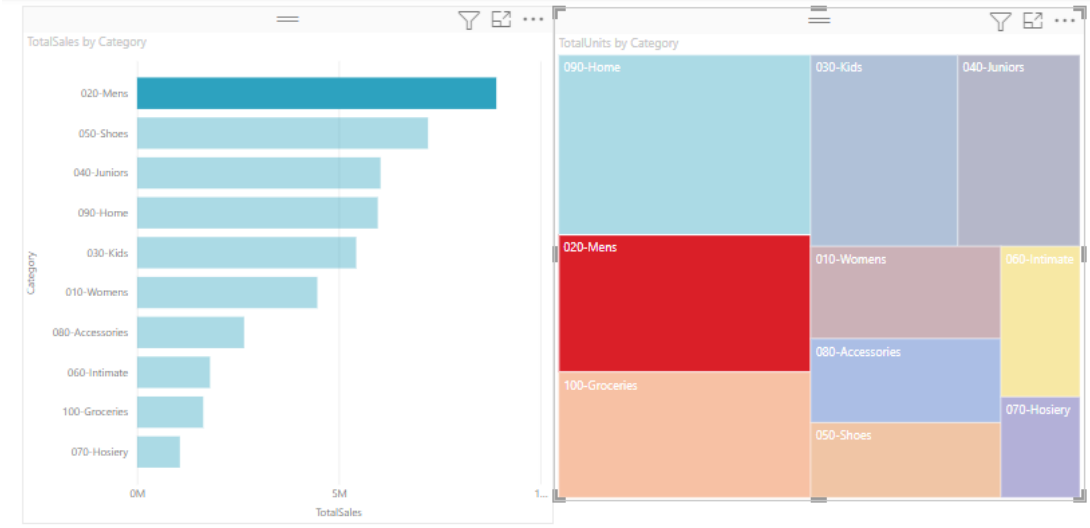


Figure 17:

070-Hosiery category's performance

Sales by Categories

Analyst also points out the performance of hosiery category. This is the category having the least revenue impact in top 10 and unit sold.

Chapter 5. Conclusion

5.1. Result and Limitations

Descriptive analytics is also useful in market research. When it comes time to glean insights from survey and focus group data, descriptive analytics can help identify relationships between variables and trends.

Concluded from Retail Analysis Dataset, we can see that Fashion Direct is the better performed chain. We also learned that 020-Mens category is the one impacting the total revenue the most in the last 2 years. Besides that, 010-Womens category is the one under-performing, this is a flag to the management to plan for production and promotion suitably. Management can also start to utilize the vendor to increase Sales per sq ft.

However, descriptive analytic itself also has the following limitations:

- Data is frozen in time: Descriptive data alone can only ever show you what happened in the past. Which limits the ability of diagnostics and predictions, this will also affect the outcome decision of a project.
- Lack of contextual database: since Descriptive analytics can only describe the data and relations, impacts. It is limited for the analysts to understand the contextual meaning of the event.

5.2. Thread development direction

We can utilize the Descriptive Analytics by understanding its roles, impacts, and importance in a Business Intelligence study. Descriptive statistics are very important because if we simply presented our raw data which would be hard to visualize what the data was showing, especially if there was a lot of it. Descriptive statistics therefore enables us to present the data in a more meaningful way, which allows simpler interpretation of the data. And based on the results

form the analysis, we can go beyond on higher levels of analytics for “the big picture” of the researched object. Which is also the greatest meaning of contemporary Data Analysis.

References

1. Graduate Programs Staff, How Do I Start a Data Project: Understanding a Project Lifecycle, August 2020 Available from: https://www.northeastern.edu/graduate/blog/data-analysis-project-lifecycle/ .
2. Terence Mills, Five Benefits Of Big Data Analytics And How Companies Can Get Started, November 2019, Available from: https://www.forbes.com/sites/forbestechcouncil/2019/11/06/five-benefits-of-big-data-analytics-and-how-companies-can-get-started/
3. Wesley Chai, What is Big Data Analytics and Why is it Important?, December 2021, Available from: https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics
4. Patrick Gibson, Types of Data Analysis, 2021, Available from: https://chartio.com/learn/data-analytics/types-of-data-analysis/
5. Jake Frankenfield, Data Analytics, September 2021, Available from: https://www.investopedia.com/terms/d/data-analytics.asp
6. Clay Halton, Predictive Analytics, June 2021, Available from: https://www.investopedia.com/terms/p/predictive-analytics.asp
7. Troy Segal, Prescriptive Analytics, July 2021, Available from: https://www.investopedia.com/terms/p/prescriptive-analytics.asp
8. Jake Frankenfield, Descriptive Analytics, December 2020, Available from: https://www.investopedia.com/terms/d/descriptive-analytics.asp
9. Hafizullah Baha, An Introduction on Descriptive Analysis: Its advantages and disadvantages, Leuphana University of Lüneburg, March 2016, page 4-5.

10. Roxy Peck, Chris Olsen, Jay L.Devore, Introduction to Statistics and Data Analysis Fifth Edition, CENGAGE Learning, 2014, Page 5-8.