Client: AdventureWorks

Consultant: Khai Ha

Course: IT420 – Business Intelligence

Date: December 2025

Executive Summary

The AdventureworkDW2022 database offered structured data that was easy to use and implemented seamlessly into the PowerBI. It provided a modernized analytical environment that was current and offered executives many data insights that would support big decisions. Over the course of 10 weeks until this full BI delivery, the work that was experienced were auditing and cleaning data, designing staging and data warehouse that completes with star schema, implementing ETL pipelines. Working with SSMS, SSIS, SSRS, and then finally Power BI to develop dashboards filled with logistics that supports predictive analytics. This portfolio will summarize the work cycle that a data analysis would go through to generate a clean dashboard for AdventureworkDW2022.

It all started with the most common yet most crucial aspect of data, the data quality and its audit. This aspect of data analysis as it is the core of any kind of connection between tables. The audit reveals missing data, wrong formatting, and many other issues such as duplications. All these possible issues expose the portfolio to risks including mismatch joins, and erroneous aggregation of the data. It shows how important it is to clean the data through repeatable ETL. That's where the job of the Data Analysis begins, cleaning the data with ETL.

Week 1- Data Quality Assessment & Source Evaluation

Deliverables: Data profiling worksheets, messy dataset assessment

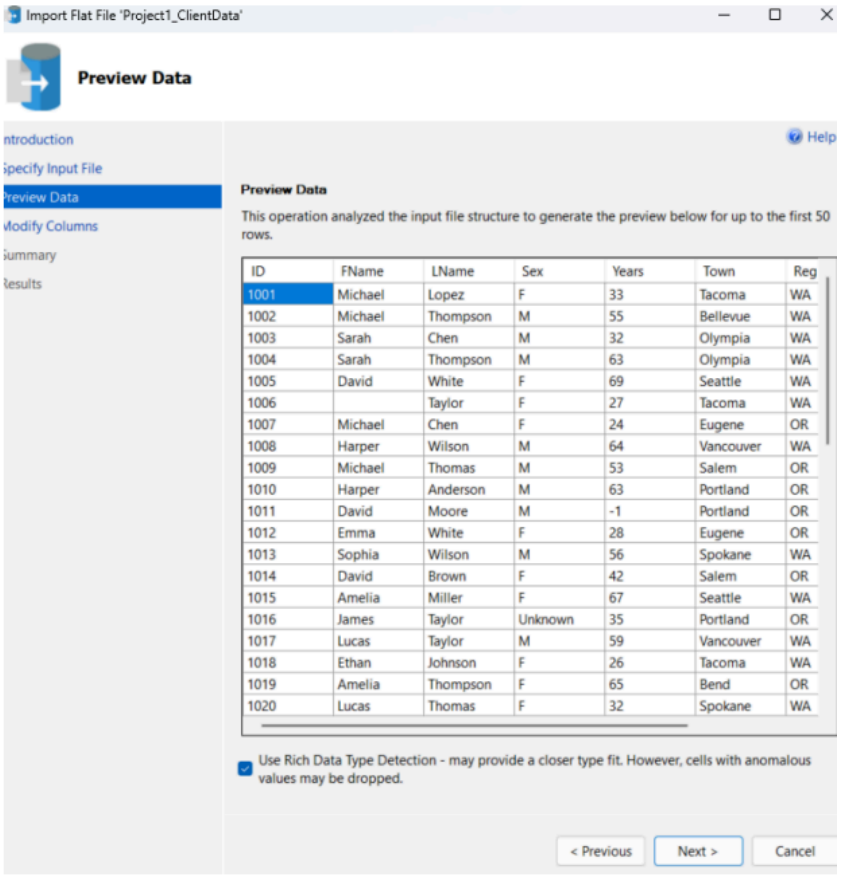
Ha_Khai_Project1

Key Actions:

In Week 1, I did a full data audit to ensure the quality of the data being provided. The files were of the messy Customer, Product, and Sales flat files. Going through the process the files source were evaluated as reliable, structurally consistent, and data integrity issues all over. The Customer file has missing data, invalid age, unknown gender values, and misspells cities. The Sales file has misaligned dates, negative quantities, invalid placeholder IDs (e.g., CustomerID 99999), and non-numeric values in TotalAmount. The Product file contains missing suppliers, missing prices, and duplicated rows. SQL server detected many issues during importation attempts, with improper data typings. This first assessment was the start of ETL cleanup for later, catching errors to prevent misinformation for reporting.

Evidence:

	CustomerID	FirstName	LastName	Gender	Age	City	State	Segment
2	1001	Michael	Lopez	F	33	Tacoma	WA	Retail
3	1002	Michael	Thompson	M	55	Bellevue	WA	Retail
4	1003	Sarah	Chen	M	32	Olympia	WA	Online
5	1004	Sarah	Thompson	M	63	Olympia	WA	Wholesale
6	1005	David	White	F	69	Seattle	WA	Retail
7	1006		Taylor	F	27	Tacoma	WA	Wholesale
8	1007	Michael	Chen	F	24	Eugene	OR	Wholesale
9	1008	Harper	Wilson	M	64	Vancouver	WA	Online
10	1009	Michael	Thomas	M	53	Salem	OR	Online
11	1010	Harper	Anderson	M	63	Portland	OR	Retail
12	1011	David	Moore	M	-1	Portland	OR	Wholesale
13	1012	Emma	White	F	28	Eugene	OR	Wholesale
14	1013	Sophia	Wilson	M	56	Spokane	WA	Online
15	1014	David	Brown	F	42	Salem	OR	Online
16	1015	Amelia	Miller	F	67	Seattle	WA	Retail
17	1016	James	Taylor	Unknown	35	Portland	OR	Retail
18	1017	Lucas	Taylor	M	59	Vancouver	WA	Wholesale
19	1018	Ethan	Johnson	F	26	Tacoma	WA	Online
20	1019	Amelia	Thompson	F	65	Bend	OR	Wholesale
21	1020	Lucas	Thomas	F	32	Spokane	WA	Online
22	1021	Charlotte	Chen	M	25	Sea_ttle	WA	Online
23	1022	Daniel	Jackson	M	42	Bellevue	WA	Online
24	1023	Ethan	Martin	F	53	Seattle	WA	Online
25	1024	Michael	Thompson	F	67	Eugene	OR	Retail
26	1025	John	Jackson	M	47	Seattle	WA	Online



Week 2 - ETL Foundations & Data Cleaning

Deliverables: SSIS ETL package, cleaned CSV outputs

Ha_Khai_Project2Reflection

Key Actions:

SSIS to build a complete ETL pipeline that filters out and corrects the bad data. The software allows the data analyst to filter out and correct missing data with mock tempt values so SSMS would accept the import. Invalid gender entries by converting “O” to “Missing Value,” corrected city names through suffix parsing and proper capitalization, and normalized all State abbreviations. Any inconsistencies and spelling issues were caught and managed through SSIS “Missing Segment.” The SSIS flows used derived columns, conditional logic, data conversions, and sequential transformations to produce fully cleaned outputs ready for staging.

Evidence:

1	CustomerID	FirstName	LastName	Gender	Age	City	State	Segment
2	1	Jacob	Sellers	F	60	Mariaberg	Wshngtn	Home Office
3	2	Joshua	Johnson	M	40	Johnsonfurt	Wshngtn	Home Office
4	3	Jesse	Berger	O	41	Lake Ashlee	Cliforna	Corporate
5	4	Amanda	Mann	O	69	Fullerland	CA	Consumer
6	5	Diane	Rasmussen	F	58	East Suzanne	Cliforna	Consumer
7	6	Dawn	Wells	F	40	Lisashire	CA	Corporate
8	7	Mark	Hanson	F	70	Carterfort	Cliforna	Home Office
9	8	Krystal	Taylor	F	60	South Jilltown	WA	Home Office
10	9	Amanda	Gordon	O	68	Hillhaven	CA	Home Office
11	10	Christina	Thornton	F	42	East James	OR	Corporate
12	11	Jesse	Ramos	O	38	Lake Ryanville		Consumer
13	12	Lisa	Stone	F	73	Mooreport	Wshngtn	Comsumer
14	13	Janice	Willis	M	69	Christophermouth	OR	Corporate
15	14	Phillip	Davis	F	37	Deanchester	OR	Corporate
16	15	Mark	Scott	M	59	Porterview	CA	Home Office
17	16	Dustin	Cardenas	O	53	Lake Kristinastad	OR	Corporate
18	17	Amy	Rodriguez	F	69	South Keith	WA	Corporate
19	18	Hannah	Webb	O	42	Port Jonathan	Wshngtn	Corporate

Process of Fixing the Issues

C...	FirstName	LastName	Gender	A...	City	State	Segment
1	Jacob	Sellers	F	60	Mariaberg	Wshngtn	Home Office
2	Joshua	Johnson	M	40	Johnsonfurt	Wshngtn	Home Office
3	Jesse	Berger	O	41	Lake Ashlee	Clforna	Corporate
4	Amanda	Mann	O	69	Fullerland	CA	Consumer
5	Diane	Rasmussen	F	58	East Suzanne	Clforna	Consumer
6	Dawn	Wells	F	40	Lisashire	CA	Corporate
7	Mark	Hanson	F	70	Carterfort	Clforna	Home Office
8	Krystal	Taylor	F	60	South Jilltown	WA	Home Office
9	Amanda	Gordon	O	68	Hillhaven	CA	Home Office
10	Christina	Thornton	F	42	East James	OR	Corporate
11	Jesse	Ramos	O	38	Lake Ryanville		Consumer
12	Lisa	Stone	F	73	Mooreport	Wshngtn	Comsumer
13	Janice	Willis	M	69	Christophermouth	OR	Corporate
14	Phillip	Davis	F	37	Deanchester	OR	Corporate
15	Mark	Scott	M	59	Porterview	CA	Home Office
16	Dustin	Cardenas	O	53	Lake Kristinastad	OR	Corporate
17	Amy	Rodriguez	F	69	South Keith	WA	Corporate
18	Hannah	Webb	O	42	Port Jonathan	Wshngtn	Corporate
19	Mark	Rodriguez	F	66	Port Sarah		Corporate
20	Mary	Davis	O	53	Lake Annette	WA	Home Office
21	Madison	Harper	O	73	East Marcus	CA	Comsumer
22	Katie	Woods	F	48	New Edwardport	CA	Corporate
23	Travis	Arroyo	M	28	West Williamview		

Week 3 - Warehouse Schema Build & ETL Expansion

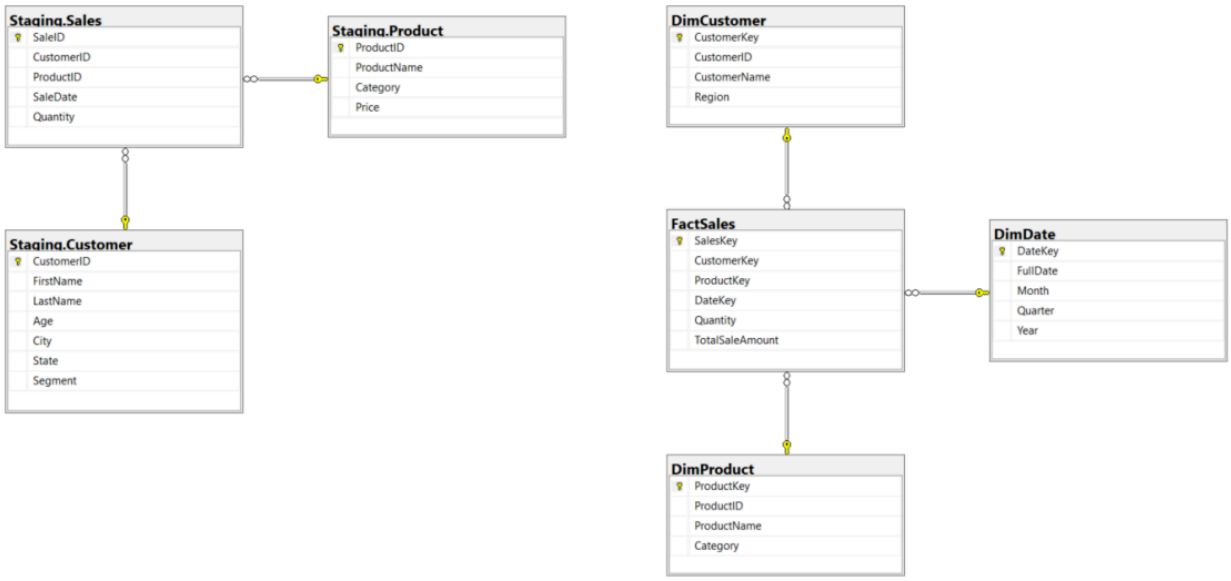
Deliverables: Dimensional model (DimCustomer, DimProduct, DimDate, FactSales), extended SSIS pipeline

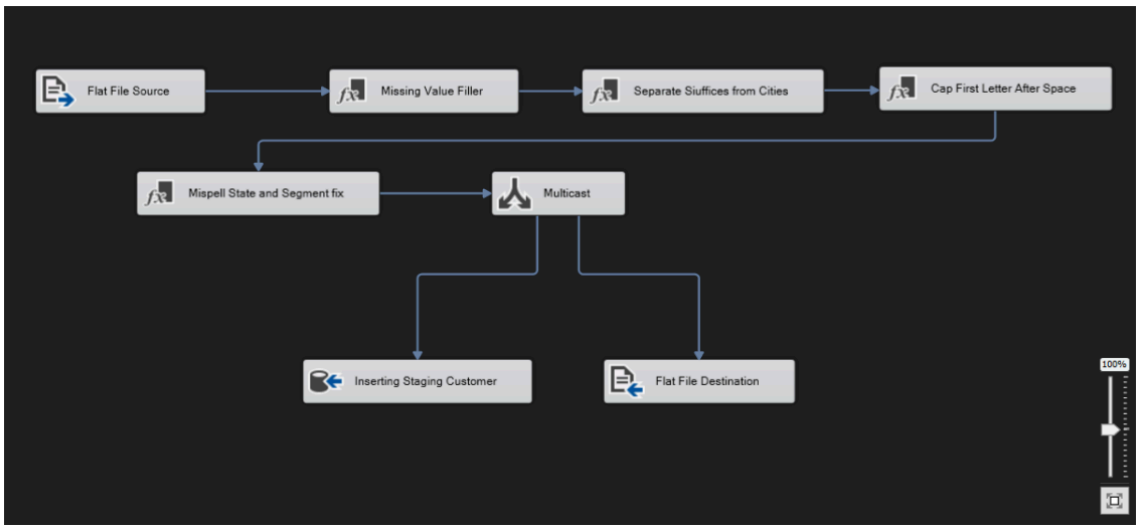
Ha_Khai_Project3

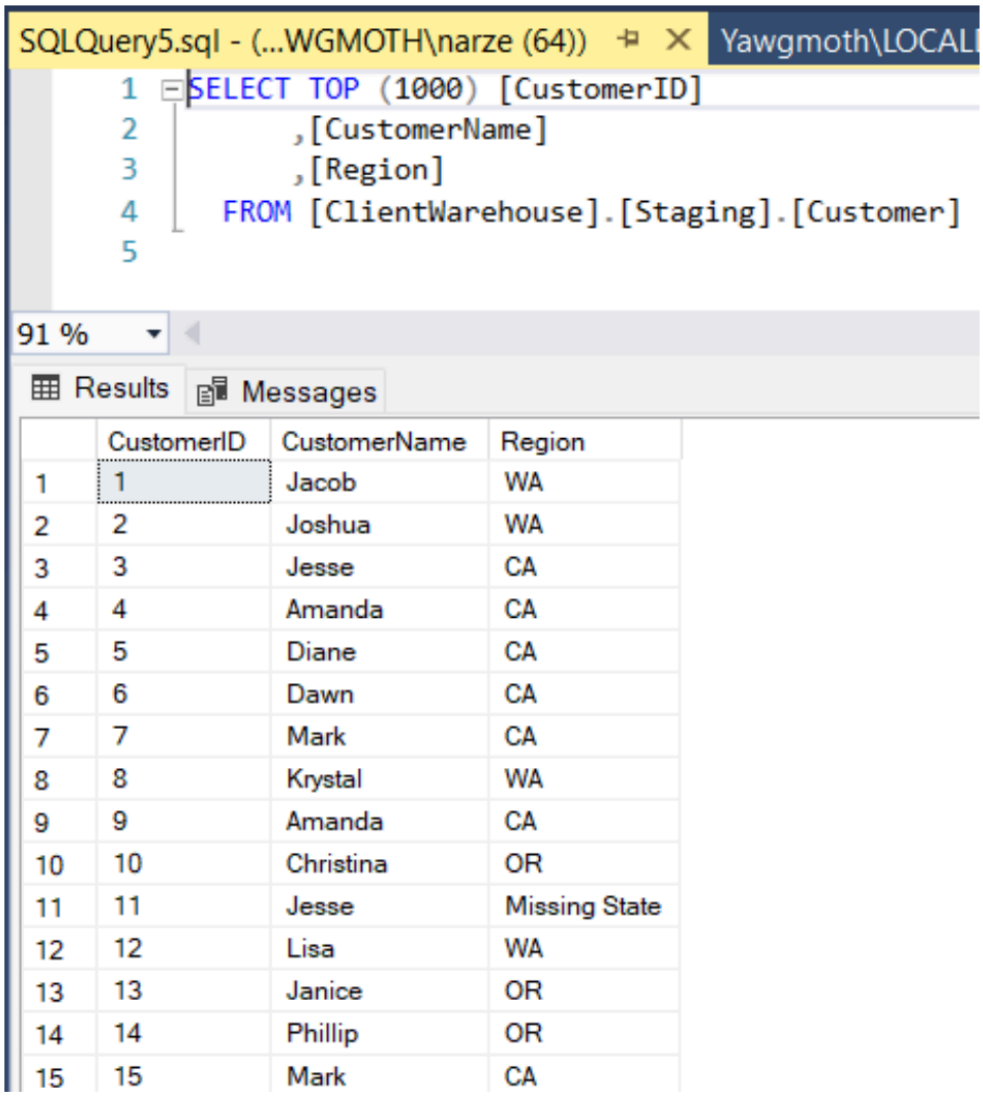
Key Actions:

Warehouse star schema was created by expanding the SSIS package from load staging into final dim/fact tables. With the creation of the ClientWarehouse database, planning of staging structure and final dim and fact table structure begin. Leading to the creation of DimCustomer, DimProduct, DimDate, and FactSales using surrogate keys, constraints, and business rules of Quantity ≥ 0. SSIS's pipeline was extended with Multi cast components to ensure clean CSVs while inserting into SSMS. SSIS software made it easier to handle data conversion without restructuring work. The lookup feature of SSIS give the data analysis an easy way to catch and filter out duplication without causing more issues. It looks into the requested table, analysis provides filter parameters, any duplication of data is filtered out before insertion.

Evidence:


Staging.Sales
SaleID
CustomerID
ProductID
SaleDate
Quantity
Staging.Product
ProductID
ProductName
Category
Price
Staging.Customer
CustomerID
FirstName
LastName
Age
City
State
Segment
DimCustomer
CustomerKey
CustomerID
CustomerName
Region
FactSales
SalesKey
CustomerKey
ProductKey
DateKey
Quantity
TotalSaleAmount
DimDate
DateKey
FullDate
Month
Quarter
Year
DimProduct
ProductKey
ProductID
ProductName
Category


Flat File Source
Missing Value Filler
Separate Siuffices from Cities
Cap First Letter After Space
Mispell State and Segment fix
Multicast
Inserting Staging Customer
Flat File Destination
100%



	CustomerID	CustomerName	Region
1	1	Jacob	WA
2	2	Joshua	WA
3	3	Jesse	CA
4	4	Amanda	CA
5	5	Diane	CA
6	6	Dawn	CA
7	7	Mark	CA
8	8	Krystal	WA
9	9	Amanda	CA
10	10	Christina	OR
11	11	Jesse	Missing State
12	12	Lisa	WA
13	13	Janice	OR
14	14	Phillip	OR
15	15	Mark	CA

Week 4 - OLTP vs. OLAP Performance Comparison

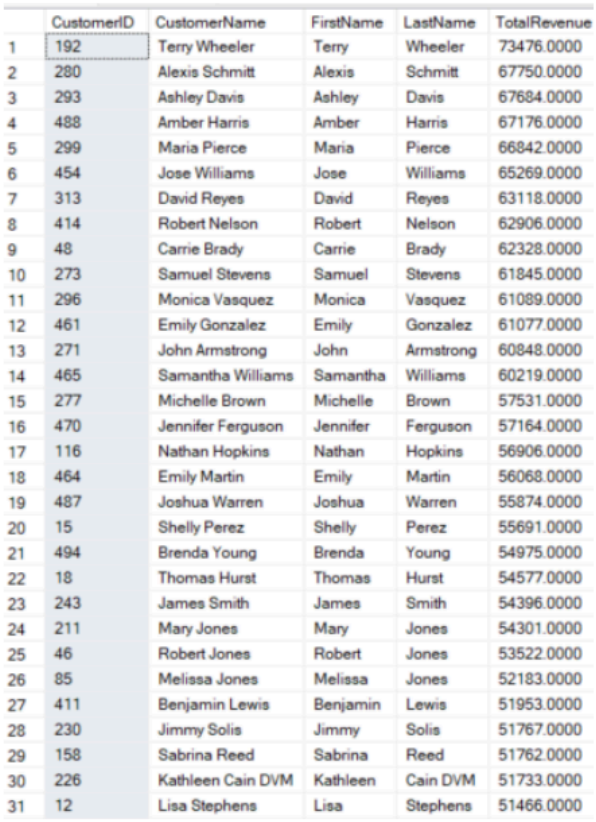
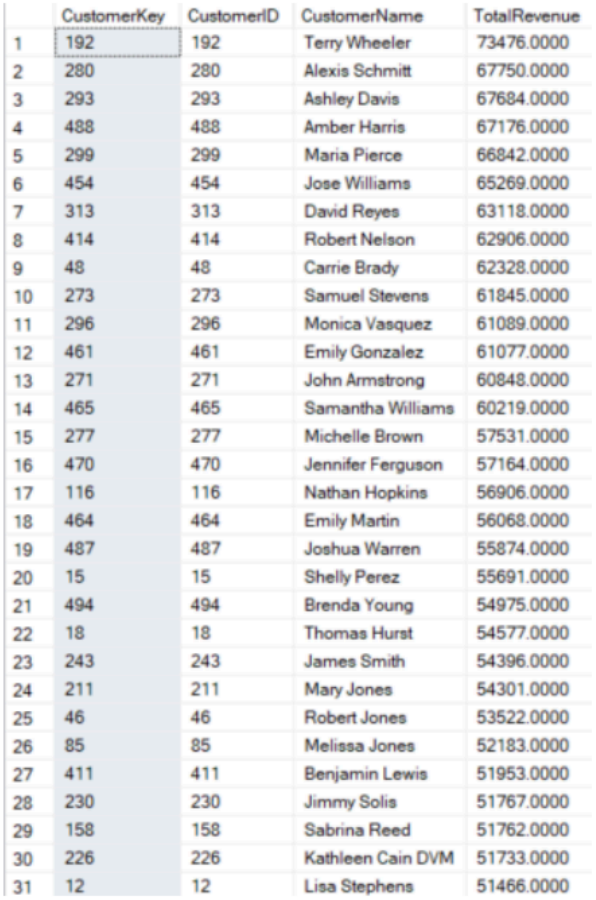
Deliverables: OLTP vs OLAP comparison report

IT420 Wk4 proj4 report

Key Actions:

Week 4 OLTP staging and OLAP dimensional table, shows the importance of interaction within the system is not meant for executive reporting. The whole process going from one to the other is messy and fill with walls and walls of data. It's a big factor that hinders speed for executives when they want fast, easy info to look through and present. Tests were created to show the difference in speed of OLTP and OLAP. It shows that OLAP returns results much faster, and works better mid-meeting presentations as queries and reports can be built in minutes without causing major issues for the technical infrastructure of the business. This is due to the difference between OLTP and OLAP as they are made and built for different things. OLTP was created with real time transaction integrity, while OLAP was optimized for aggregation, business intelligence, and fast scaning. OLAP for BI to ensure reliable, scalable performance.

Evidence:

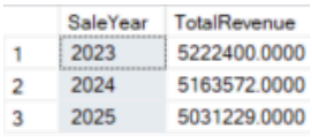
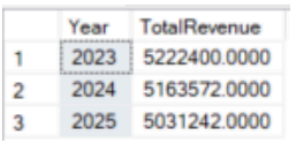
OLTP(Staging)	OLAP(Dim)
	
(500 rows affected) **Completion time:** **2025-10-29T13:39:33.3746756-07:00**	**(500 rows affected)** **Completion time:** **2025-10-29T14:03:12.5215882-07:00**

OLTP(Staging)	OLAP(Dim)

	State	TotalRevenue
1	IA	577503.0000
2	CA	494928.0000
3	WA	468226.0000
4	ME	445960.0000
5	KY	444148.0000
6	HI	439485.0000
7	MS	434126.0000
8	DC	433714.0000
9	MD	427794.0000
10	NH	416664.0000
11	MI	415589.0000
12	IL	409649.0000
13	OH	407675.0000
14	NJ	402411.0000
15	OK	373122.0000
16	NY	345261.0000
17	OR	342042.0000
18	CT	335865.0000
19	TX	332371.0000
20	ID	331143.0000
21	DE	329299.0000
22	NC	320381.0000
23	GA	319224.0000
24	VT	317433.0000
25	SC	315819.0000
26	KS	313361.0000
27	AZ	312461.0000
28	AK	294267.0000
29	AL	293095.0000
30	IN	292302.0000
31	PA	270617.0000
32	WI	266716.0000
33	ND	260691.0000
34	LA	239045.0000
35	NM	235603.0000
36	NE	229682.0000
37	WY	222997.0000
38	MN	222628.0000

	State	TotalRevenue
1	IA	577503.0000
2	CA	494928.0000
3	WA	468226.0000
4	ME	445960.0000
5	KY	444161.0000
6	HI	439485.0000
7	MS	434126.0000
8	DC	433714.0000
9	MD	427794.0000
10	NH	416664.0000
11	MI	415589.0000
12	IL	409649.0000
13	OH	407675.0000
14	NJ	402411.0000
15	OK	373122.0000
16	NY	345261.0000
17	OR	342042.0000
18	CT	335865.0000
19	TX	332371.0000
20	ID	331143.0000
21	DE	329299.0000
22	NC	320381.0000
23	GA	319224.0000
24	VT	317433.0000
25	SC	315819.0000
26	KS	313361.0000
27	AZ	312461.0000
28	AK	294267.0000
29	AL	293095.0000
30	IN	292302.0000
31	PA	270617.0000
32	WI	266716.0000
33	ND	260691.0000
34	LA	239045.0000
35	NM	235603.0000
36	NE	229682.0000
37	WY	222997.0000
38	MN	222628.0000

(51 rows affected) **Completion time:** **2025-10-29T13:41:19.2720454-07:00**	**(51 rows affected)** **Completion time:** **2025-10-29T14:05:16.4024461-07:00**

OLTP(Staging)	OLAP(Dim)
	
(3 rows affected) **Completion time: 2025-10-29T13:45:49.0167427-07:00**	**(3 rows affected)** **Completion time: 2025-10-29T14:07:03.6918948-07:00**

	SaleYear	TotalRevenue
1	2023	5222400.0000
2	2024	5163572.0000
3	2025	5031229.0000

	Year	TotalRevenue
1	2023	5222400.0000
2	2024	5163572.0000
3	2025	5031242.0000

Week 5 - SSRS Static Reporting (Tabular & Chart Reports)

Deliverables: SSRS Report – Total Revenue by Category, Top 10 Products by Revenue

IT420 Week5 Project SSRS Report…

Key Actions:

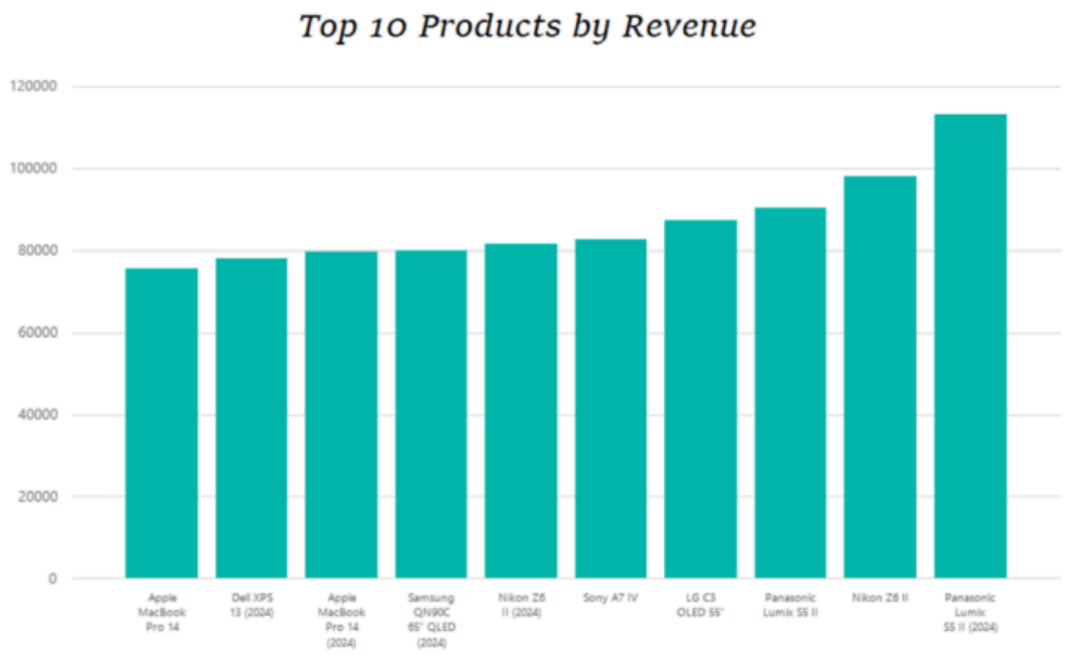
Two SSRS reports were made for executives in a tabular report summarizing Total Revenue by Category and a bar chart showing Top 10 Products by Revenue. The category report showed Electronics as highest performing category, while the Top 10 chart revealed Panasonic as a leading brand in multiple products. This information provides the executives clear information that they can use to make big decisions that can influence how the business does in the future.

Evidence:

Executive_Report_Tabular

Category	Total Qty	Total Revenue
Electronics	2798	2550314.23
Home Decor	2987	912393.07
Cosmetic	2102	38047.98
Health	2609	30622.51
Groceries	2957	16599.21

Total Revenue by Category



Top 10 Products by Revenue Chart

Week 6 - SSRS Interactive Reporting Enhancements

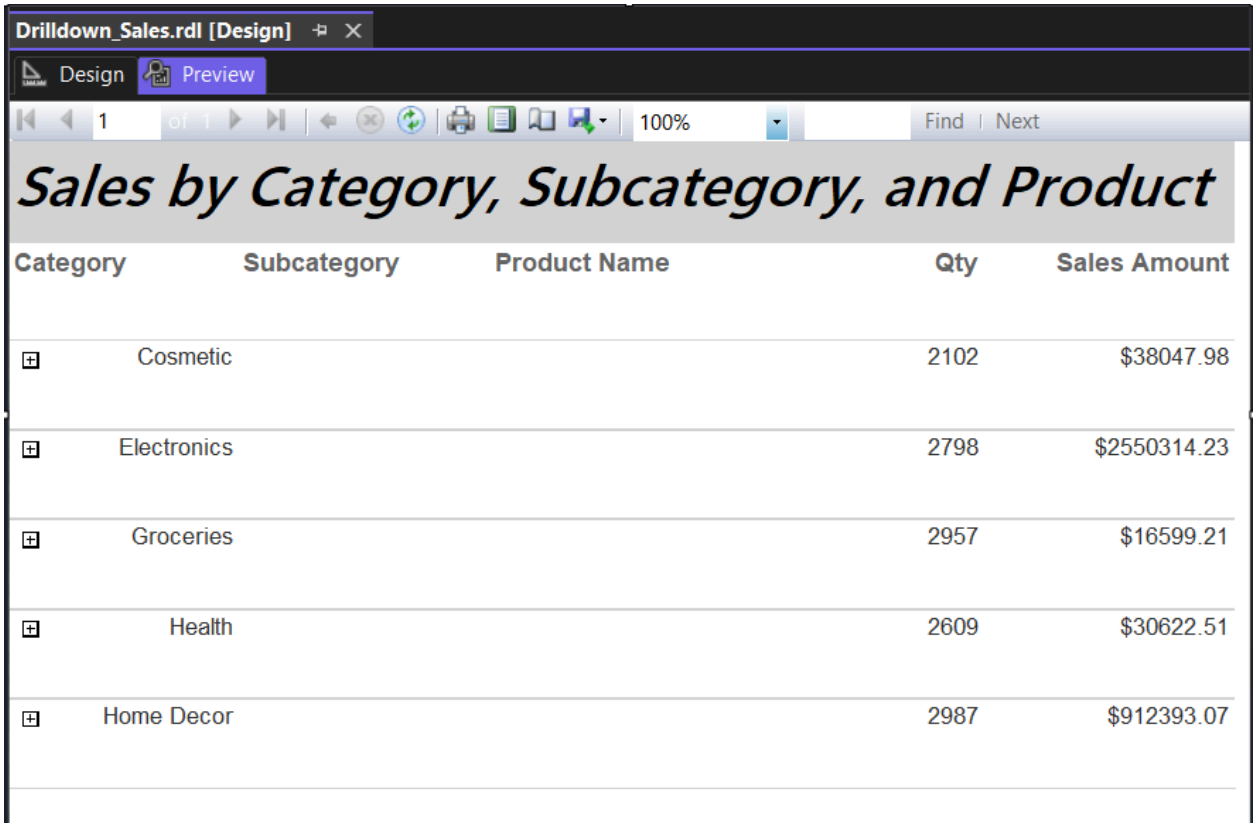
Deliverables: Interactive SSRS report with drill-down + date filters

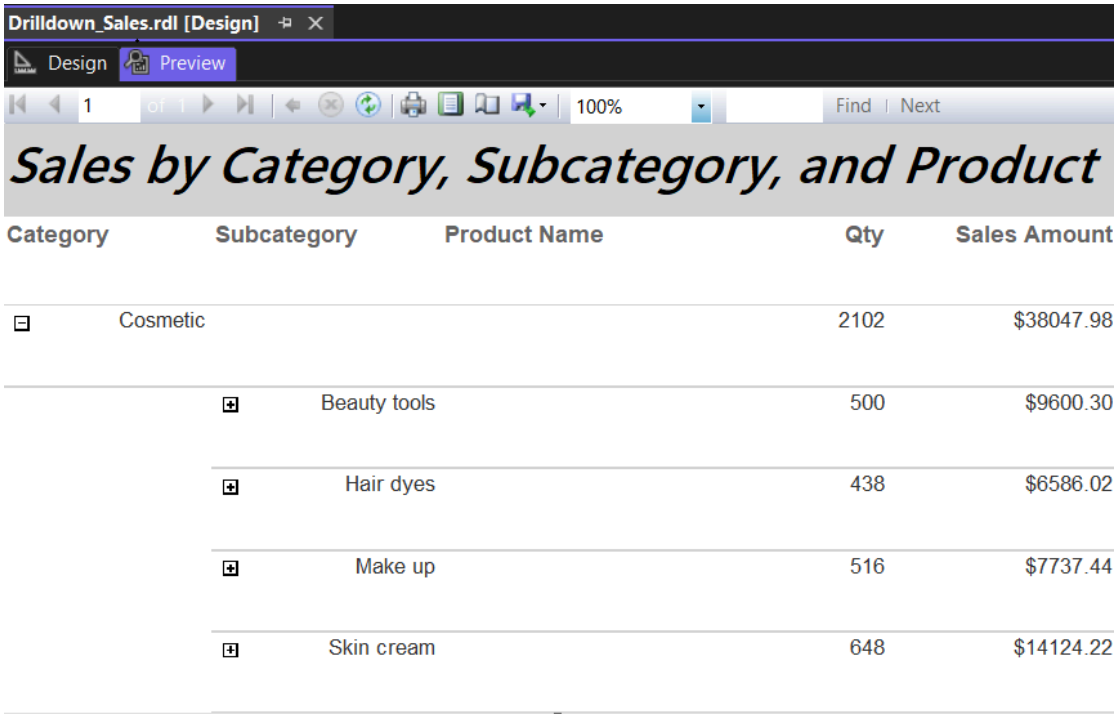
IT420 Wk 6 Project SSRS Reflect…

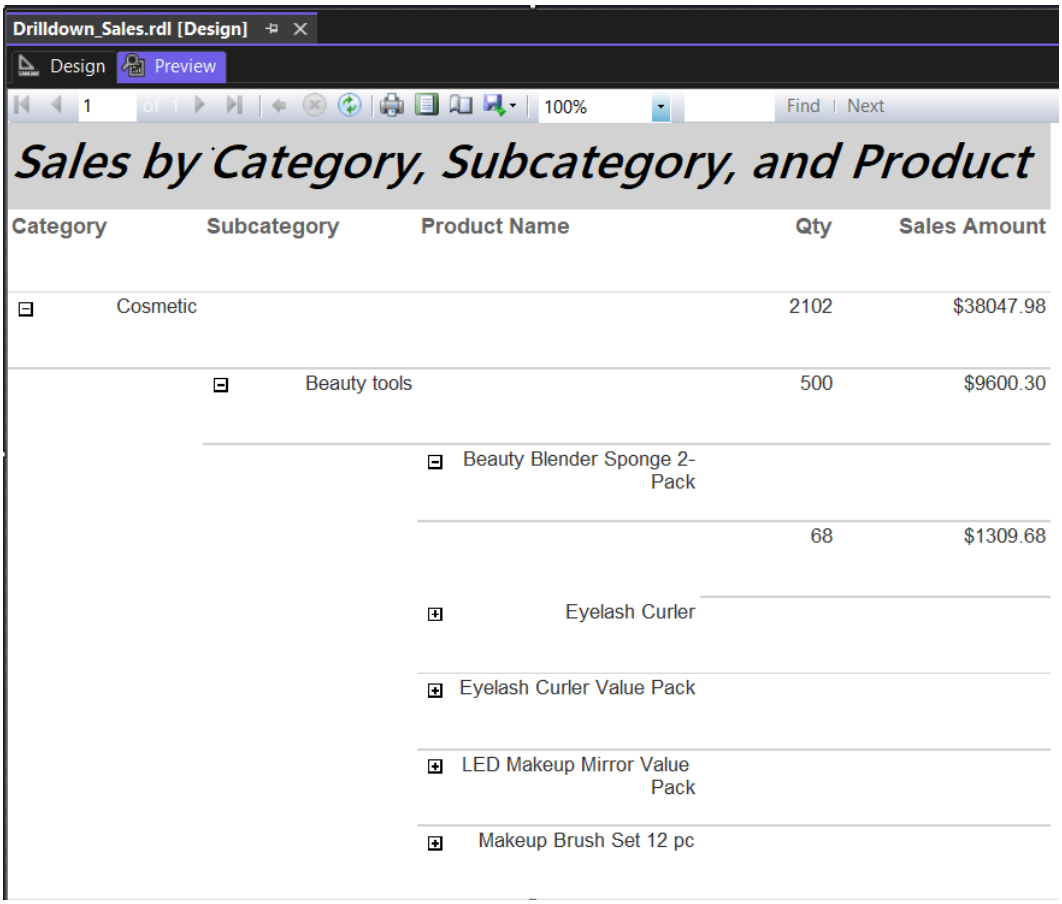
Key Actions:

The SSRS also offers drill-down interactions and a date-range parameter that allows executives to explore categories and products scope to what they want to see, making the report more user friendly and visually less clutter. This makes it so that a data analyst won't have to constantly generate reports as requested, they can provide features and notes for executives to scope into what they need to present their point.

Evidence:




Drilldown_Sales.rdl [Design]
Design
Preview
1
of 1
100%
Find | Next
Sales by Category, Subcategory, and Product
Category
Subcategory
Product Name
Qty
Sales Amount
Cosmetic
2102
$38047.98
Beauty tools
500
$9600.30
Hair dyes
438
$6586.02
Make up
516
$7737.44
Skin cream
648
$14124.22


Drilldown_Sales.rdl [Design]
Design
Preview
1
of 1
100%
Find | Next
Sales by Category, Subcategory, and Product
Category
Subcategory
Product Name
Qty
Sales Amount
Cosmetic
2102
$38047.98
Beauty tools
500
$9600.30
Beauty Blender Sponge 2-Pack
68
$1309.68
Eyelash Curler
Eyelash Curler Value Pack
LED Makeup Mirror Value Pack
Makeup Brush Set 12 pc

Week 7 - Power BI Executive Dashboard

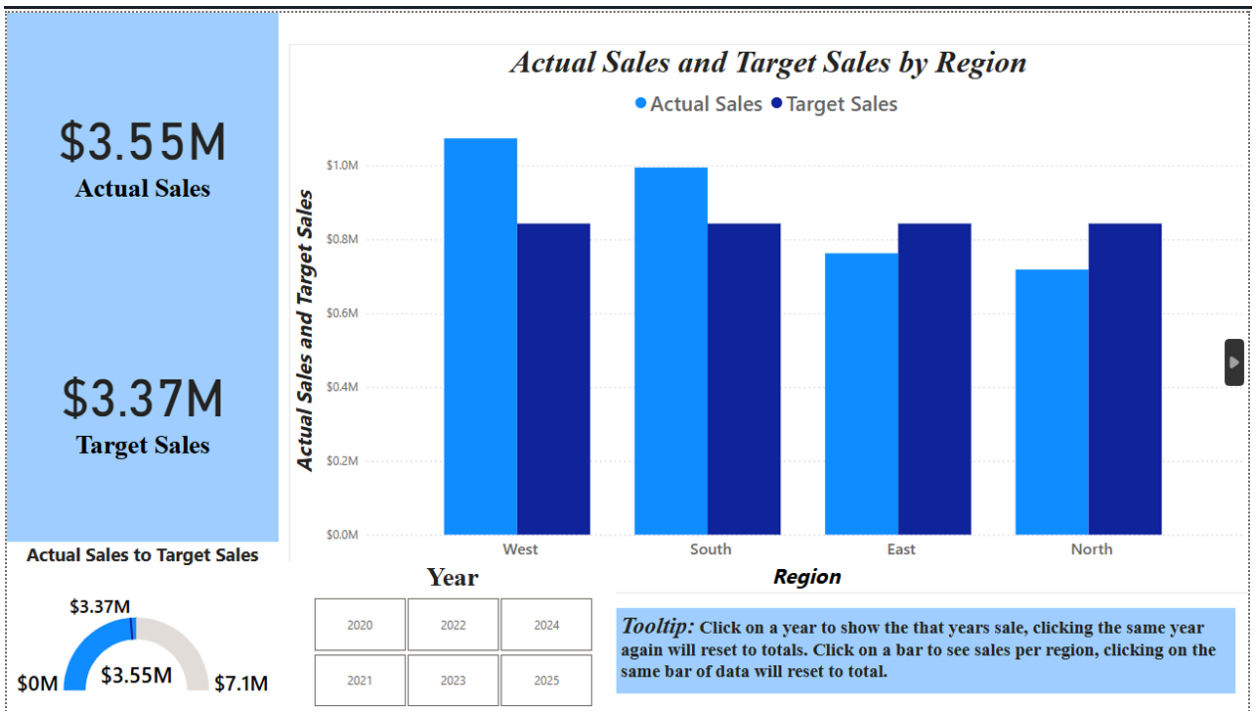
Deliverables: Multi-visual executive dashboard with slicers and interactions

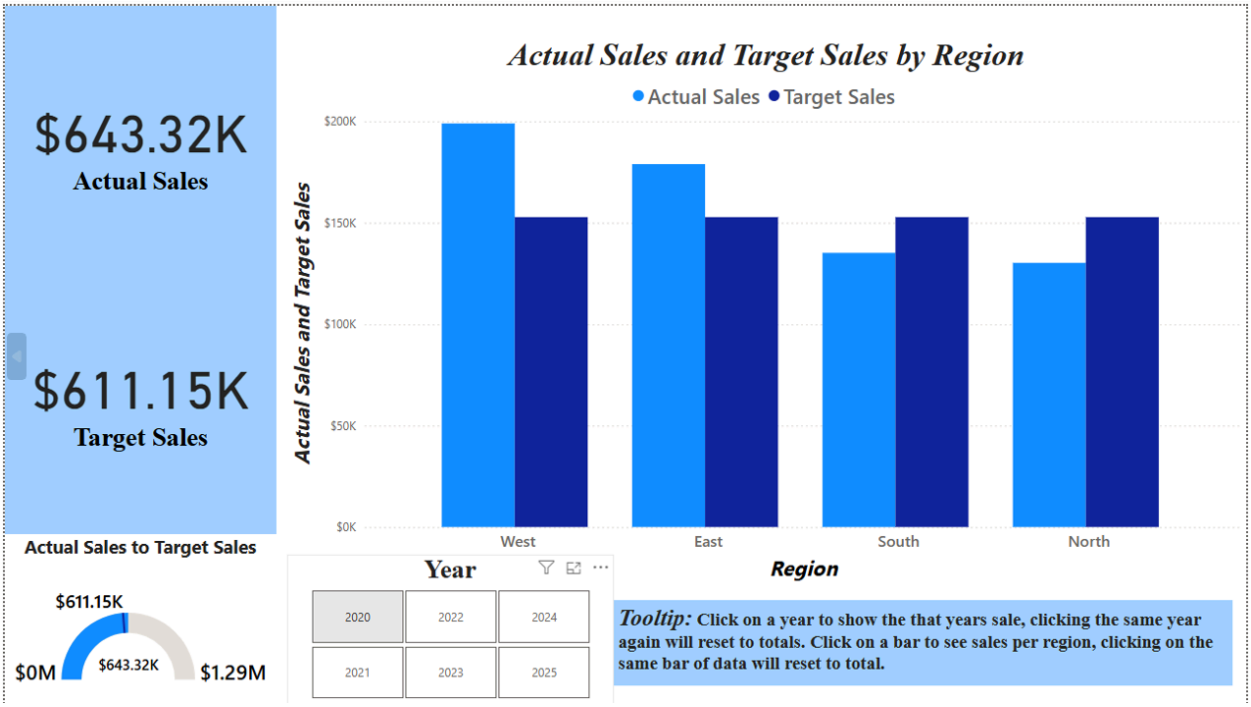
IT420 Wk 7 Project Power BI Exe…

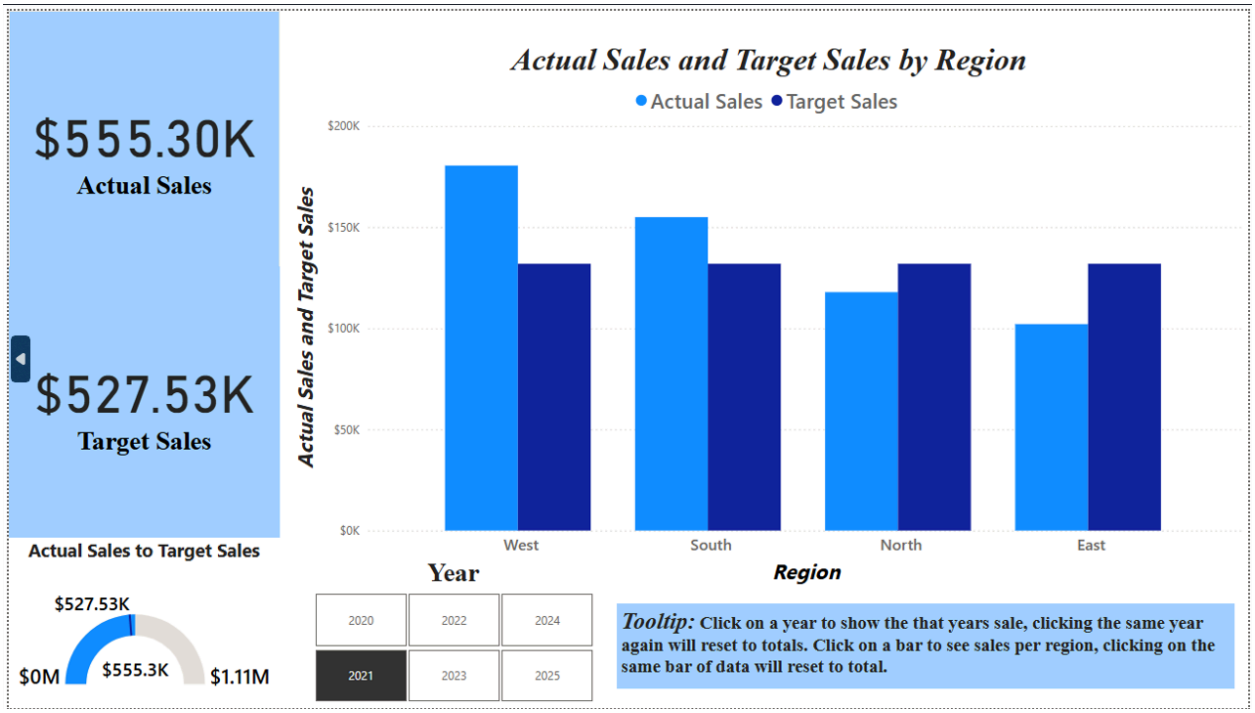
Key Actions:

Week 7 Power BI dashboard featuring KPI cards for Actual and Target Sales are finally in use, a gauge that visualizes performance toward the target, and a clustered bar chart comparing actual vs. target sales by region. A year slicer gives the option for fast time-based filtering, and all visuals are dynamically changing to adjust and show what's being selected. A tooltip was added to guide users through the interactive dashboards. These features make the whole dashboard cleaner and ready to use, making non data analysts a data expert themselves.

Evidence:




$643.32K
Actual Sales
$611.15K
Target Sales
Actual Sales to Target Sales
$611.15K
$0M
$643.32K
$1.29M
Actual Sales and Target Sales by Region
Actual Sales
Target Sales
$200K
$150K
$100K
$50K
$0K
Actual Sales and Target Sales
West
East
South
North
Region
Year
2020
2022
2024
2021
2023
2025
Tooltip: Click on a year to show the that years sale, clicking the same year again will reset to totals. Click on a bar to see sales per region, clicking on the same bar of data will reset to total.


$555.30K
Actual Sales
$527.53K
Target Sales
Actual Sales to Target Sales
$527.53K
$0M
$555.3K
$1.11M
Actual Sales and Target Sales by Region
Actual Sales
Target Sales
$200K
$150K
$100K
$50K
$0K
Actual Sales and Target Sales
West
South
North
East
Region
Year
2020
2022
2024
2021
2023
2025
Tooltip: Click on a year to show the that years sale, clicking the same year again will reset to totals. Click on a bar to see sales per region, clicking on the same bar of data will reset to total.

Week 8 - Storytelling Dashboard

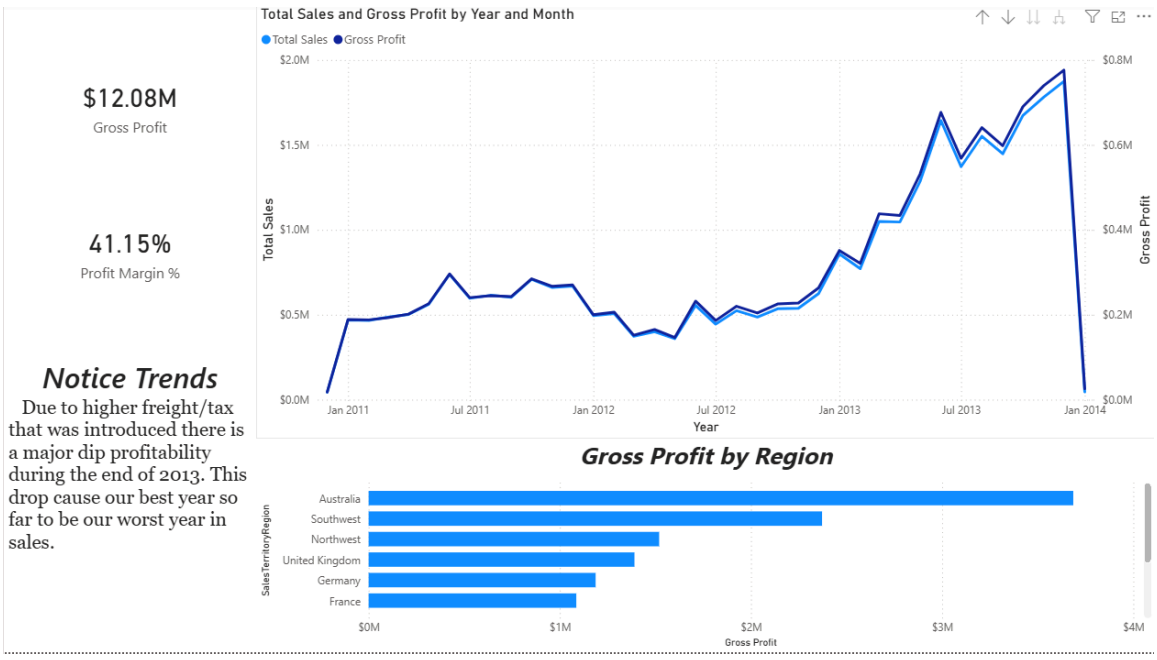
Deliverables: Power BI storytelling dashboard

IT420 Wk8 Project Reflection

Key Actions:

Week 8, the storytelling-driven dashboard emphasizing profitability, one of the most important financial indicators for real-world businesses. The dashboard includes Gross Profit and Profit Margin KPI cards, a centered line chart showing Total Sales by month across years, and a clustered bar chart displaying Gross Profit by country. Visual design choices, such as contrasting but harmonious line colors, centered titles and legends, and a prominently placed annotation help guide the narrative. The annotation describes a mock factor explaining a sales dip in late 2013, demonstrating how contextual storytelling can enrich executive insights.

Evidence:



Week 9 - Executive Decision Support Dashboard (Multi-Page)

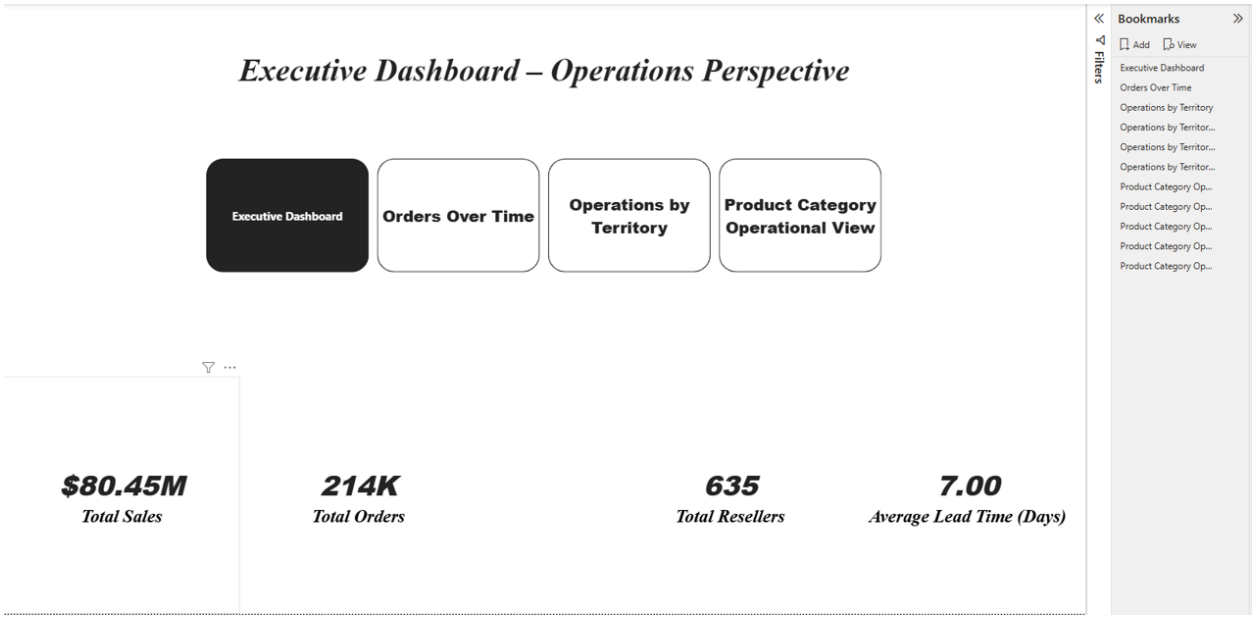
Deliverables: Power BI multi-page executive dashboard

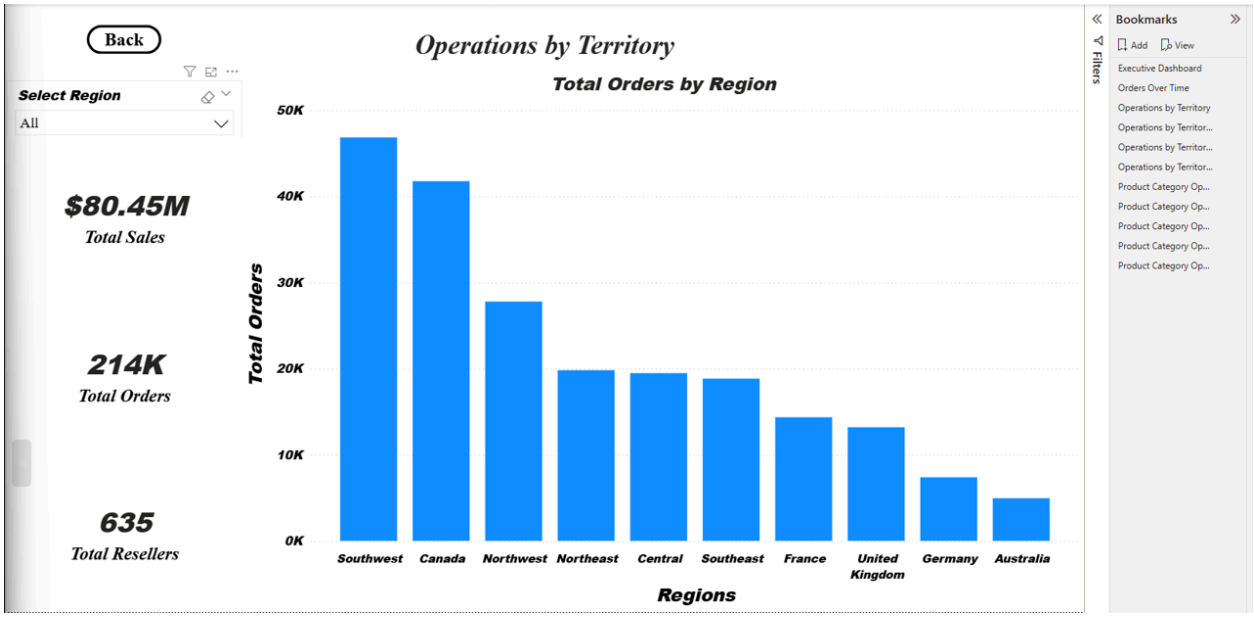
IT420 Wk 9 Project Executive De…

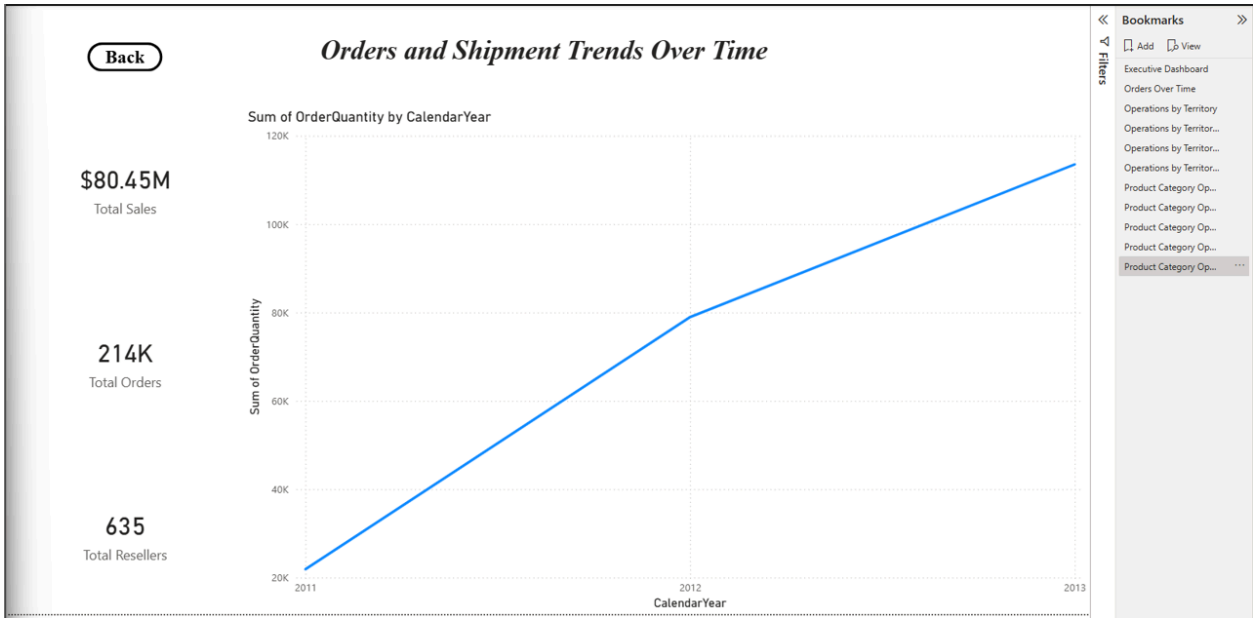
Key Actions:

Week 9 expanded the dashboard into a multi-page Executive Decision Support system featuring four executive-level KPIs(Total Sales, Total Orders, Total Resellers, and Average Lead Time). The homepage serves as a navigation hub, linking to pages such as Orders Over Time, Operations by Territory, and Product Category Performance. Orders Over Time reveals yearly performance trends. Operations by Territory highlights regional strengths and weaknesses, Product Category insights indicate which products to expand or discontinue. These are valuable information as these can help executives negotiate with imports, and supply chains.

Evidence:


Executive Dashboard – Operations Perspective
Executive Dashboard
Orders Over Time
Operations by Territory
Product Category Operational View
$80.45M
Total Sales
214K
Total Orders
635
Total Resellers
7.00
Average Lead Time (Days)
Bookmarks
Add
View
Executive Dashboard
Orders Over Time
Operations by Territory
Operations by Territor...
Operations by Territor...
Operations by Territor...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Filters


Back
Operations by Territory
Select Region
All
$80.45M
Total Sales
214K
Total Orders
635
Total Resellers
Total Orders by Region
50K
40K
30K
20K
10K
0K
Total Orders
Southwest
Canada
Northwest
Northeast
Central
Southeast
France
United Kingdom
Germany
Australia
Regions
Bookmarks
Add
View
Executive Dashboard
Orders Over Time
Operations by Territory
Operations by Territor...
Operations by Territor...
Operations by Territor...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Filters


Back
Orders and Shipment Trends Over Time
Sum of OrderQuantity by CalendarYear
$80.45M
Total Sales
214K
Total Orders
635
Total Resellers
120K
100K
80K
60K
40K
20K
Sum of OrderQuantity
2011
2012
2013
CalendarYear
Filters
Bookmarks
Add
View
Executive Dashboard
Orders Over Time
Operations by Territory
Operations by Territor...
Operations by Territor...
Operations by Territor...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...

Back
Orders and Shipment Trends Over Time
Order Quantity Over Calendar Year
$80.45M
Total Sales
214K
Total Orders
635
Total Resellers
120K
100K
80K
60K
40K
20K
Total Order Quantity
2011
2012
2013
Calendar Year
Filters
Bookmarks
Add
View
Executive Dashboard
Orders Over Time
Operations by Territory
Operations by Territor...
Operations by Territor...
Operations by Territor...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...
Product Category Op...

Week 10 - Predictive Insights Dashboard (Clustering & Classification)

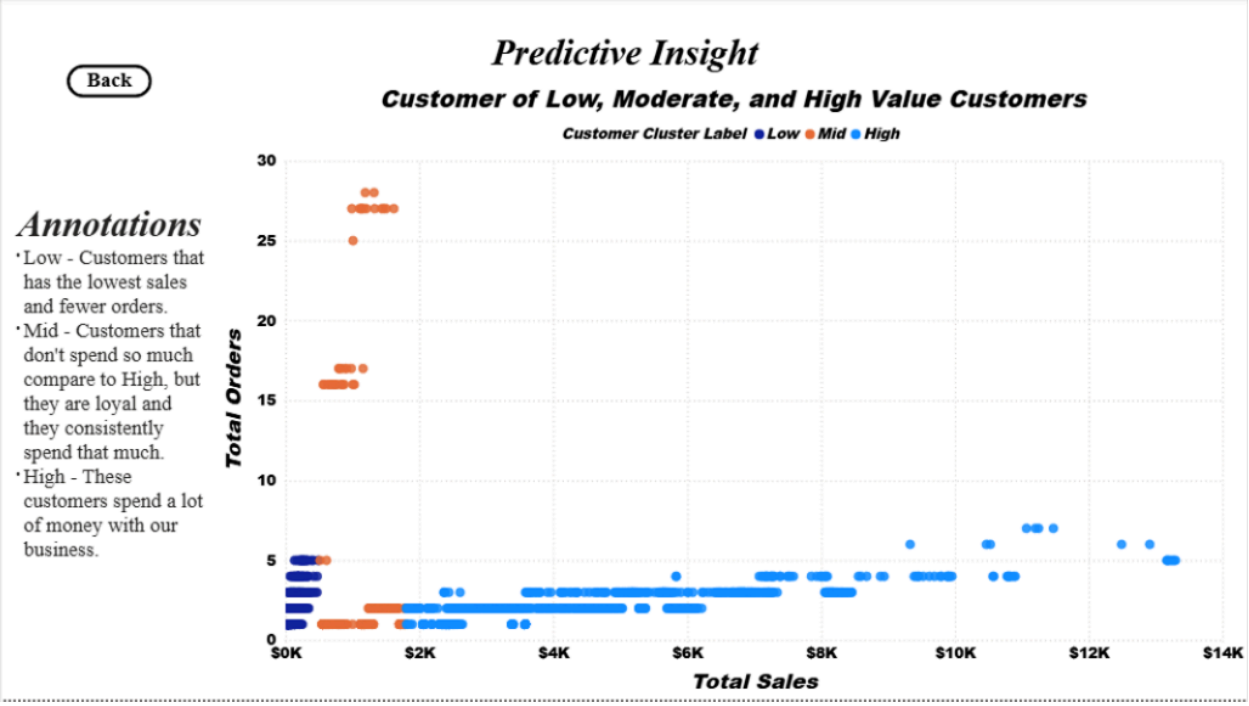
Deliverables: Predictive analytics enhancement to the Executive Dashboard

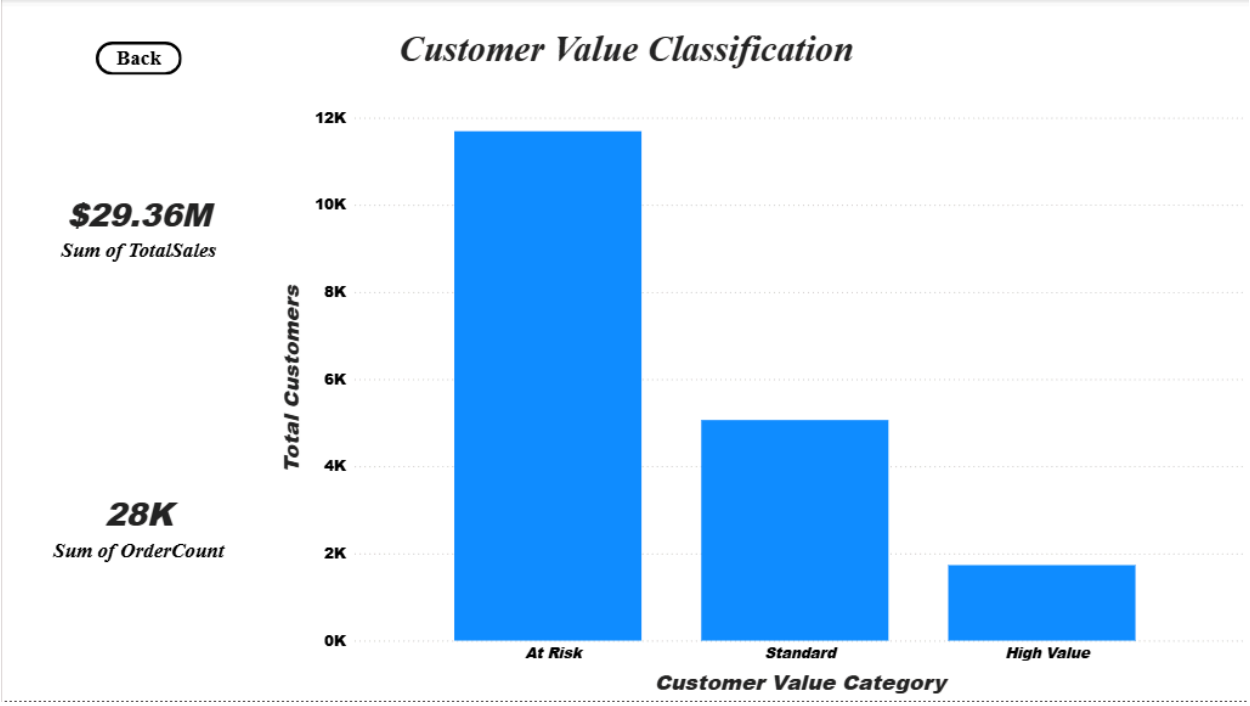
IT420 Project Wk 10 Memo

Key Actions:

Week 10 integrated predictive analytics into the dashboard using customer clustering and classification models. Clustering grouped customers into Low Value, Mid Value, and High Value segments using Total Sales and Total Orders to illustrate customer engagement and revenue contribution patterns. Classification categorized customers as At Risk, Standard, or High Value, allowing executives to quickly identify retention priorities and growth opportunities. Strategic recommendations included protecting High Value customers, re-engaging. At Risk customers due to their potential impact, and growing Standard customers through upselling and cross-selling. These predictive pages transformed the dashboard from descriptive to forward-looking, equipping executives with an analytical tool for strategic decision-making.

Evidence:


Back
Predictive Insight
Customer of Low, Moderate, and High Value Customers
Customer Cluster Label Low Mid High
Annotations
Low - Customers that has the lowest sales and fewer orders.
Mid - Customers that don't spend so much compare to High, but they are loyal and they consistently spend that much.
High - These customers spend a lot of money with our business.
Total Orders
30
25
20
15
10
5
0
$0K
$2K
$4K
$6K
$8K
$10K
$12K
$14K
Total Sales


Back
Customer Value Classification
$29.36M
Sum of TotalSales
28K
Sum of OrderCount
Total Customers
12K
10K
8K
6K
4K
2K
0K
At Risk
Standard
High Value
Customer Value Category

Final Recommendations for the Client

1. **Data Audit:** Data was analyzed, filtered and clean to ensure quality data for display.
2. **Implemented ETL pipelines:** To refresh the warehouse daily or hourly to business needs.
3. **Expand predictive analytics:** Prep to support churn prevention, demand forecasting, and inventory optimization.

   Develop standardized KPIs: to ensure consistency across departments.

   Promote easy to use interactive dashboard BI: using Power BI dashboards to reduce dependency on IT for routine reporting.

Reflection on Professional Growth

Over the course of weeks building up to this final submission, I learn to endure and be independent. In a professional setting I'll be all alone with competitors for help. It's nice in a professional setting that data analysts are paid to handle these situations.

The 10 weeks offered me time to strengthen my comfortability with SSMS, while learning new exciting software for business intelligence. The process of going from raw data to something presentable is quite simple, logic and concepts are there. The unfamiliarity with a new ui without a tutorial lead to challenges of time spent searching through live resources to just find the proper method to achieve the result required by the assignment.

Provided software such SSMS, SSIS, SSRS, and PowerBI works together very well. SSIS taught me how to design efficiency resilient ETL pipelines using lookups, derived columns, conditional splits, and data conversions. Building a star schema gives a better visual on how dimensional modeling and why OLAP structures outperform OLTP for workloads. Developing SSRS reports helped me appreciate detail and clarity, while Power BI dashboards pushed for understanding integrating visual and data into a storytelling, and user-centric design. Adding predictive insights in Week 10 expanded my toolkit into data science concepts.