Data Analytics Practicing Project

Original Data Schema

sales_data_sample(

ORDERNUMBER,

QUANTITYORDER,

PRICEEACH,

ORDERLINENUMBER,

SALES,

ORDERDATE,

STATUS,

QTE_ID,

MONTH_ID,

YEAR_ID,

PRODUCTLINE,

MSRP,

PRODUCTCODE,

CUSTOMERNAME,

PHONE,

ADDRESSLINE1,

ADDRESSLINE2,

CITY,

STATE,

POSTALCODE,

COUNTRY,

TERRITORY,

CONTACTLASTNAME,

CONTACTFIRSTNAME,

DEALSIZE

)

Data Description

The dataset represents a sample of a company’s sales transactions from **6th January 2003** to **31st March 2005**. It contains detailed information about orders, products, customers, and geographic locations.

The data include:

- **Order Information:** Includes Order ID, Order Date, Order Line ID, Order Status, Quantity, Price, and Sales amount.
- **Product Information:** Product Line, Product ID, MSRP, and Deal Size.
- **Customer Information:** Customer Name, Contact Names, Phone, Address, City, State, Postal Code, Country, and Territory.

- **Date Dimensions:** Month ID, Quarter ID, Year ID.
- **Geographical Information:** City, State, Postal Code, Country, and Territory.

Observed Data Issues Before Preprocessing:

1. Missing values in columns such as Territory, State, City, Postal Code, and Phone.
2. Inconsistent formats in Phone numbers, Postal Codes, and abbreviations in State names.
3. Duplicate entries in Customers, Products, and Date tables.
4. Mixed or unclear naming conventions (e.g., NY vs NYC, NA vs North America).

Preprocessing Operations

1. **Column Renaming** – standardized names for readability and consistency:

Original	New
ORDERNUMBER	Order ID
QUANTITYORDER	Ordered Quantity
PRICEEACH	Each Quantity Price
ORDERLINENUMBER	Order Line ID
SALES	Sales
ORDERDATE	Order Date
STATUS	Order Status
QTR_ID	Quarter ID
MONTH_ID	Month ID
YEAR_ID	Year ID
PRODUCTLINE	Product Line
PRODUCTCODE	Product ID
CUSTOMERNAME	Customer Name

Original	New
PHONE	Phone
ADDRESSLINE1	Address Line 1
ADDRESSLINE2	Address Line 2
CITY	City
STATE	State
POSTALCODE	Postal Code
COUNTRY	Country
TERRITORY	Territory
CONTACTLASTNAME	Contact Last Name
CONTACTFIRSTNAME	Contact First Name
DEALSIZE	Deal Size

1. **Territory Replacement:**
    - `NA` → `North America`
2. **Phone Number Cleanup:**
    - Remove spaces, dashes, and parentheses to standardize format.
3. **State Replacement:**

Original	Replacement
BC	British Columbia
CA	California
CT	Connecticut
MA	Massachusetts
NH	New Hampshire
NSW	New South Wales
NV	Nevada
NY	New York
PA	Pennsylvania

- Missing values → `unknown`

1. **Postal Code Preprocessing:**
    - Remove extraneous characters (e.g., `()`,).
    - Replace blank values with `00000`.
2. **City Standardization:**
    - Replace `NYC` → `New York City`.
3. **Normalization & Table Creation:**

Table	Columns
Sales	Order ID **(PK)**, Ordered Quantity, Each Quantity Price, Order Line ID, Sales, Calculated Sales, Order Date, Order Status, Product ID **(FK)**, Customer Name **(FK)**
Customers	Customer Name **(PK)**, Phone, Address Line 1, Address Line 2, City, State, Postal Code, Country, Territory, Contact Last Name, Contact First Name
Products	Product Line, MSRP, Product ID **(PK)**, Deal Size
Date	Order ID, Quarter ID, Month ID, Year ID, Total Sales
Location	State **(PK)**, Longitude, Latitude

- **Duplicate Removal:**
    - Customers[Customer Name]
    - Products[Product ID]
    - Date[Order Date]

Data Modeling

- **Relationships:**
    - Sales → Products (many-to-one)
    - Sales → Customers (many-to-one)
    - Customers → Location (many-to-one)
- **Star Schema Structure:**
    - Central fact table: **Sales**

- Dimension tables: Customers, Products, Date, Location

New Table Added

Table: Namemonthly_sales_forcast1

Fields: ORDERDATE, SALES

Description: Predicted sales for the next year using Python

Pages & Visualization

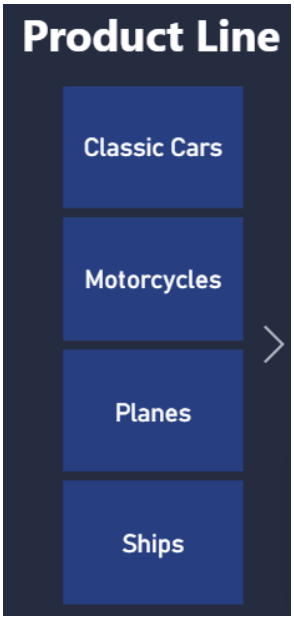
Page 1: Sales Performance

Slicers

- **Year Slicer:** Filters the page by selected year.

- **State:** Filters the page by selected state.

- **Product Line:** Filters the page by selected product line.



KPIs Cards

1. **Totals Orders**

   Purpose:

   - Monitor overall business activity.
   - Quickly identify periods with high or low order volumes.
   - Serve as a reference for other KPIs, such as Average Deal Size or Total Sales.

   Calculation:

   - DAX formula

     ```
     Total Orders = DISTINCTCOUNT(Sales[Order ID])
     ```

2. **Sales YoY%**

   Purpose:

   Shows the percentage change in total sales compared to the same period in the previous year.

   Calculation:

   - DAX formula

     ```
     OrdersDITINCTCOUNTOrder IDSales YoY% =
     IF(
        ISFILTERED('Sales'[Order Date]),
     ```

```
    ERROR("Time intelligence quick measures can only be grouped or filter
ed by the Power BI-provided date hierarchy or primary date column."),
    VAR __PREV_YEAR =
        CALCULATE(
            SUM('Sales'[Sales]),
            DATEADD('Sales'[Order Date].[Date], -1, YEAR)
        )
    RETURN
        DIVIDE(SUM('Sales'[Sales]) - __PREV_YEAR, __PREV_YEAR)
)
```

3. **Sum of Sales**

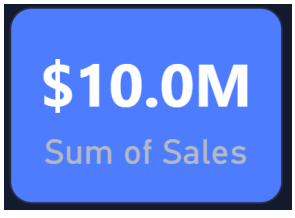
   Purpose:

   Shows the sum of all sales in the dataset.

   Calculation:

   - DAX formula

```
Total Sales = SUM(Sales[Sales])
```



4. **AVG Order Price**

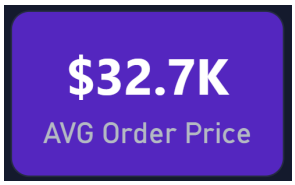
   Purpose:

   Shows the average value of each order in the dataset.

   Calculation:

   - DAX formula

```
AVG Order Price = sum(Sales[Sales])/Sales[Total Orders]
```



Area Chart

Title

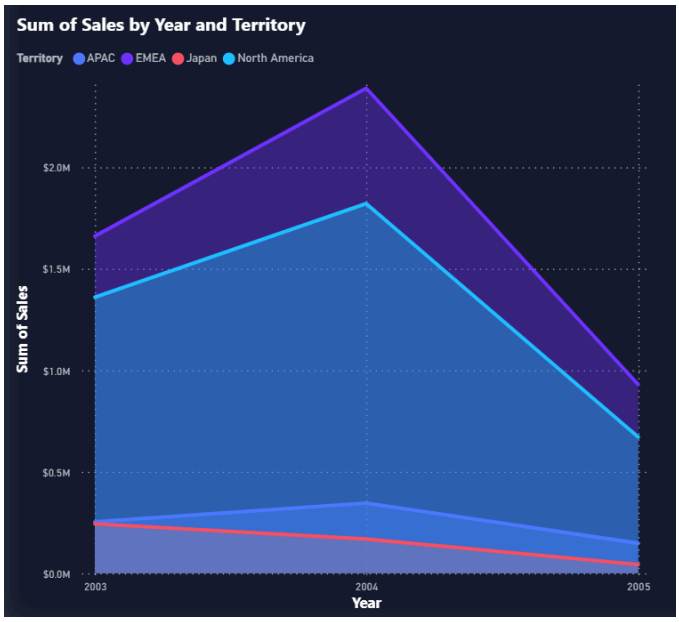
Sum of Sales by Year and Territory

Purpose

- Identify which regions generate the most sales.
- Track sales trends per territory year by year.
- Spot regions with growth or decline.

Visualization Details:

- **X-axis:** Year
- **Y-axis:** Sum of Sales
- **Legend / Color:** Territory (APAC, EMEA, Japan, North America)
- Displayed as a **stacked area chart** to show both totals and contributions of each territory.



Donut Chart

Title

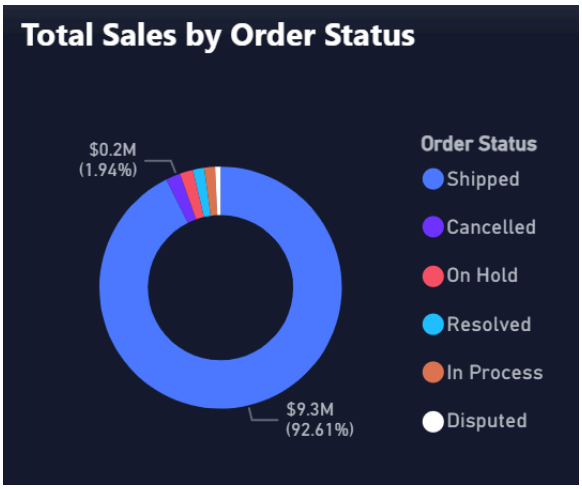
Total Sales by Order Status

Purpose

Shows how total sales are distributed across order statuses.

Visualization Details:

- **Data:** Sales grouped by Order Status
- **Legend / Color:** Order Status (Shipped, Cancelled, On Hold, Resolved, In Process, Disputed)



Funnel Chart

Title

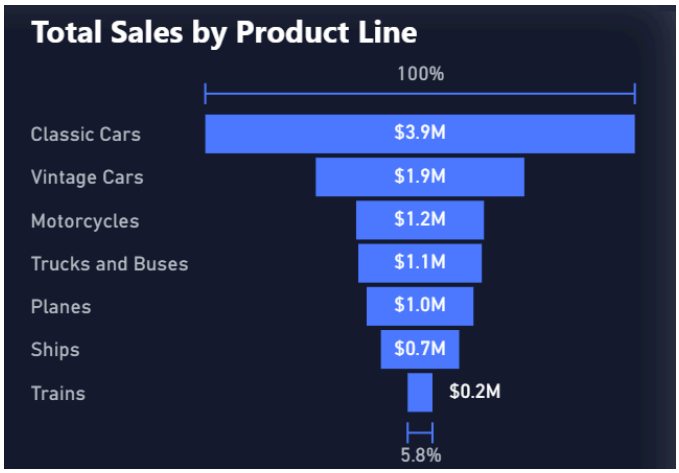
Total Sales by Product Line

Purpose

Shows total sales for each product line, highlighting which categories contribute most to revenue.

Visualization Details:

- **Data:** Sales grouped by Product Line
- **Legend / Color:** Product Line (Classic Car, Vintage Cars, Motorcycles, Trucks and Buses, Planes, Ships, Trains)



Scatter Plot

Title

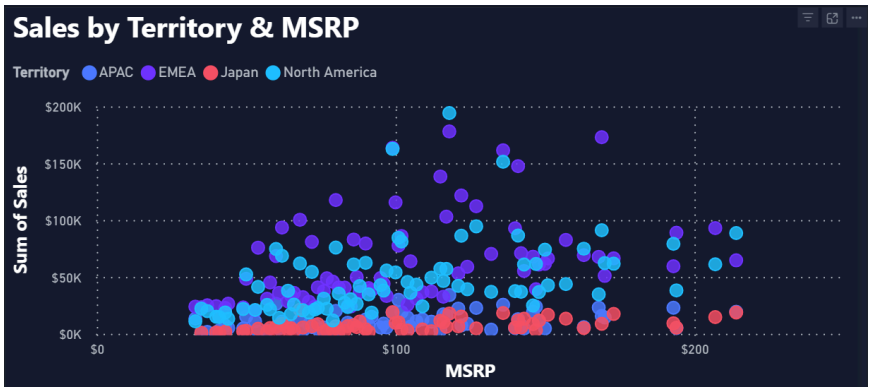
Sales by Territory & MSRP

Purpose

Shows the relationship between product price (MSRP) and total sales across different territories.

Visualization Details:

- **Data:** Sum of Sales vs MSRP by Territory
- **X-axis:** MSRP
- **Y-axis:** Sum of Sales
- **Legend / Color:** Territory (APAC, EMEA, Japan, North America)



Top Customers Table

Title

Top Customers by Sales

Purpose

Displays the highest-performing customers based on total sales, along with their rank and percentage contribution to overall sales.

Visualization Details:

- **Data:**
  - Customer Name
  - Rank **(New Measure)**

    DAX formula

```
Rank =
RANKX(
    ALL(Customers[Customer Name]),
[Total Sales],
    ,
    DESC,
```

```
    Dense
)
```

    - Sum of Sales
    - % of Total Sales
- **Visualization Type:** Table with in-cell bar indicators for Sales
- **Sorting:** Ranked from highest to lowest Sum of Sales
- **Highlights:**
    - Shows absolute sales values
    - Shows percentage contribution to total sales
    - Includes visual bar representation for quick comparison
- **Interactivity**
    - **Drill Through:** Selecting a Customer Name opens the **Customer Details** drill-through page.

Page 2: Customer Details Page

Header Card

Title

Customer Name

Purpose

Displays the name of the selected customer coming from the drill-through action.

Visualization Details:

- **Data:** Customer Name
- **Interactivity:** Driven by drill-through from the main report

KPI Cards

Total Orders

Title

Total Orders **(Measure)**

DAX formula

```
Total Orders = DISTINCTCOUNT(Sales[Order ID])
```

Purpose

Shows the total number of orders placed by the selected customer.

Visualization Details:

- **Data:** Count of Orders for the selected customer

Total Sales

Title

Total Sales **(Measure)**

DAX formula

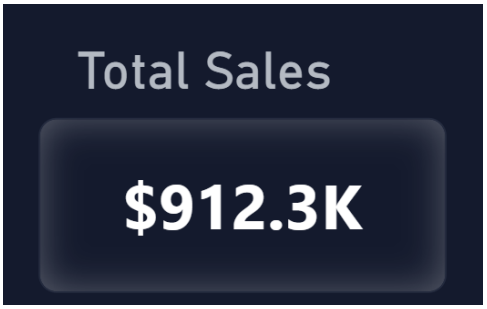
```
Total Sales = SUM(Sales[Sales])
```

Purpose

Shows the total sales amount generated by the selected customer.

Visualization Details:

- **Data:** Sum of Sales for the customer



First Order Date

Title

First Order Date **(Measure)**

DAX formula

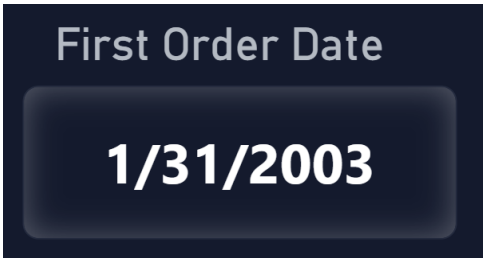
```
First Order Date = MIN(Sales[Order Date])
```

Purpose

Displays the earliest date on which the customer placed an order.

Visualization Details:

- **Data:** Minimum Order Date



Last Order Date

Title

Last Order Date **(Measure)**

DAX formula

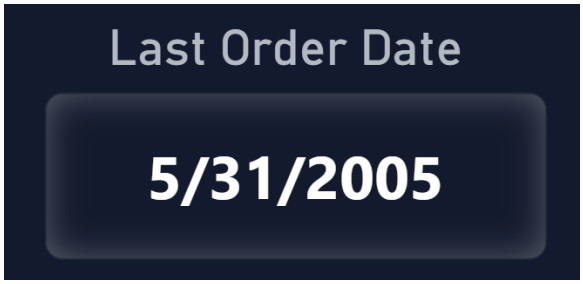
```
Last Order Date = MAX(Sales[Order Date])
```

Purpose

Displays the most recent date on which the customer placed an order.

Visualization Details:

- **Data:** Maximum Order Date



Pie Chart

Title

Sales by Category

Purpose

Shows how the customer's total sales are distributed across dynamically calculated sales categories.

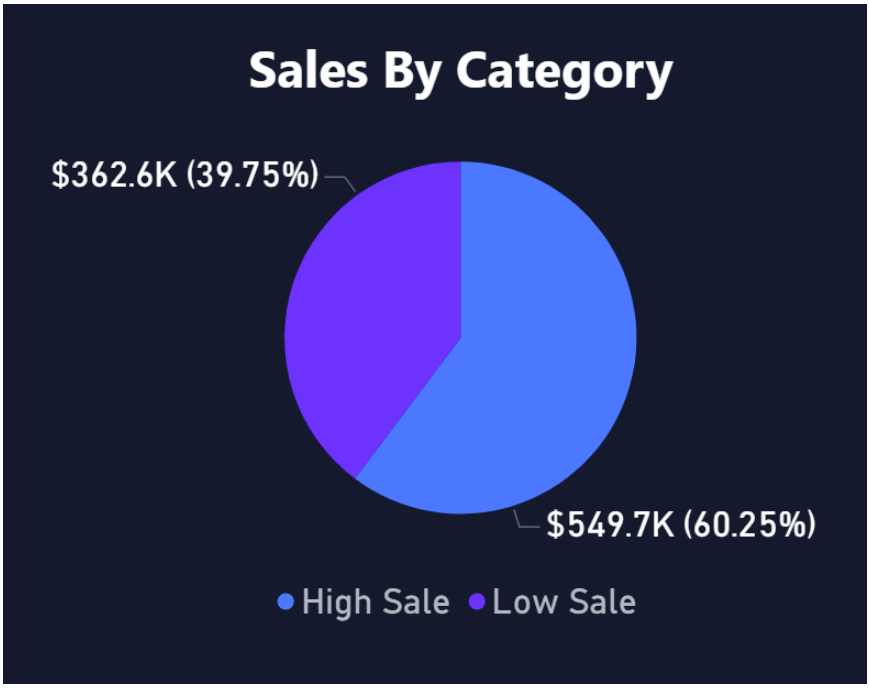
Visualization Details:

- **Data:** Sales Category **(calculated column)**

  DAX formula

```
Sales Category =
VAR AvgSale = CALCULATE(AVERAGE(Sales[Sales]), ALL(Sales))
RETURN
IF(Sales[Sales] >= AvgSale, "High Sale", "Low Sale")
```

- **Categories:** High Sale, Low Sale
- **Legend / Color:** High Sale, Low Sale



Column Chart

Title

Sum of Sales by Week Days

Purpose

Displays the distribution of the customer's sales across days of the week based on a calculated weekday column.

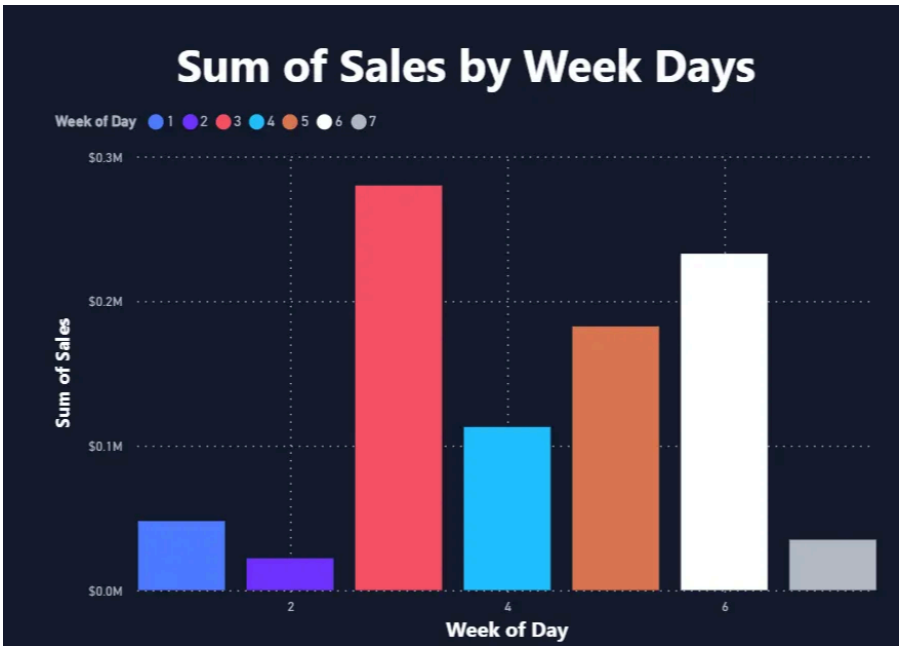
Visualization Details:

- **Data:** Sales grouped by **Week of Day**
- **Calculated Column:**

  DAX formula

  ```
  Week of Day = WEEKDAY(Sales[Order Date],1)
  ```

- **Axis:**
  - **X-Axis:** Week of Day (1–7)
  - **Y-Axis:** Sum of Sales
- **Legend / Color:** Each weekday represented with a distinct color



Slicers

Title

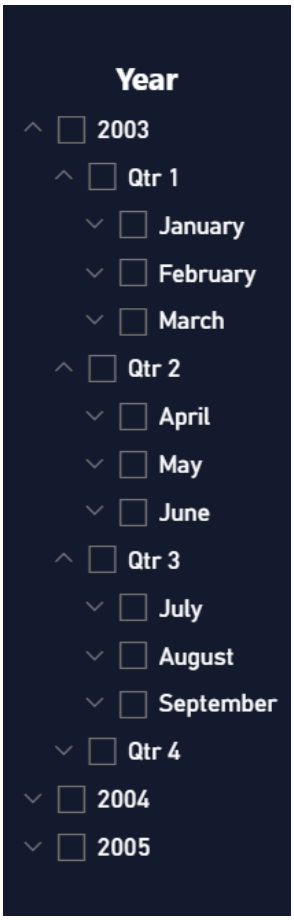
Year Filter

Purpose

Allows filtering of the customer’s sales and behavior by year, quarter, month and day.

Visualization Details:

- **Data:** Year → Quarter → Month hierarchy
- **Interactivity:** Affects all visuals on the page



Page 3: Sales Trend

Line Chart

Title

Sales Trend with Forecast

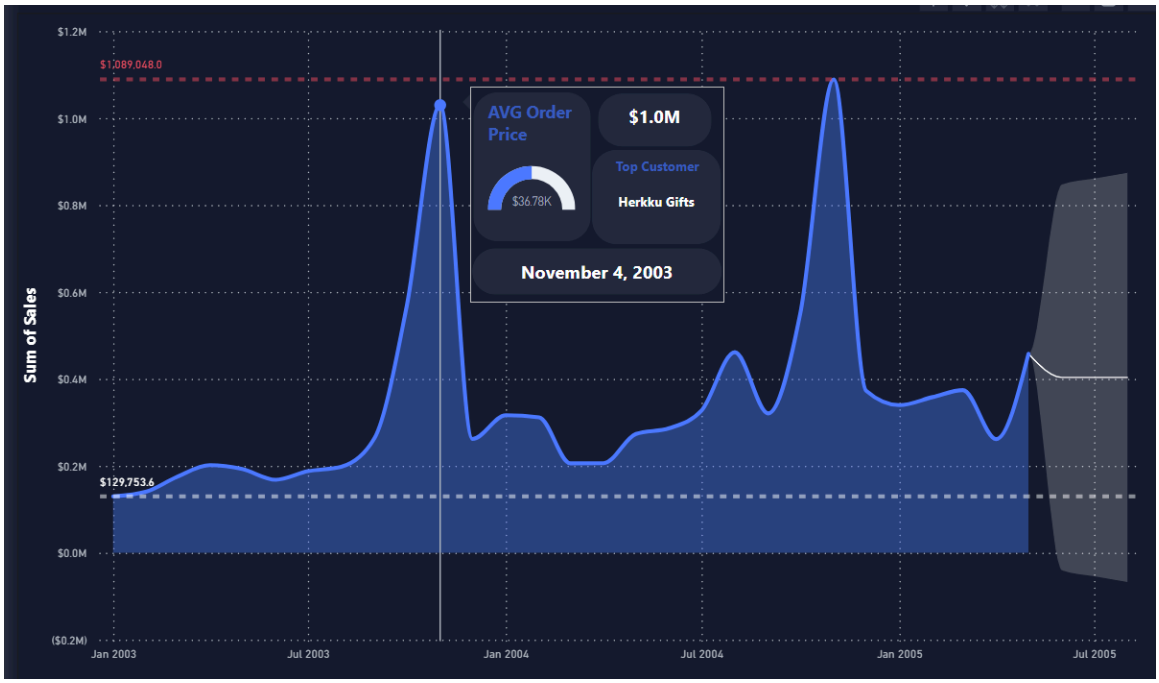
Purpose

Displays monthly sales over time, highlights minimum and maximum sales points, and forecasts the next 3 months.

Visualization Details:

- **Data:** Sales by Month and Year
- **Forecast:** Next 3 months of projected sales included
- **Reference Lines:**
  - **Max Sales: Red** dotted line
  - **Min Sales: White** dotted line
- **Axis:**
  - **X-Axis:** Month & Year
  - **Y-Axis:** Sum of Sales
- **Visualization Type:** Line chart with forecast and reference lines

- Connected with Sales Trend Tooltip



Sales Trend Tooltip

Title

Sales Trend Tooltip

Purpose

Provides additional information when hovering over points in the Sales Trend line chart, giving deeper insight into sales behavior for the selected month.

Visualization Details:

- **AVG Order Price:** Displays the average order value for the hovered month
- **Total Sales:** Shows the monthly total sales amount
- **Top Customer:** Identifies the customer with the highest sales for that month
- **Top Order Date:** Displays the specific date with the highest order value during that month

Interactivity:

- Appears when hovering over any data point on the Sales Trend line chart

- Contextual values update based on the selected month

Narrative

Title

Sales Trend Narrative

Purpose

Provides an automatically generated written explanation of the sales trend, highlighting key insights such as increases, decreases, fluctuations, and forecast behavior.

Visualization Details:

- **Data Source:** Same data used in the Sales Trend line chart (Sales by Month & Year + Forecast)
- **Content:** Auto-generated narrative summarizing:
  - Overall trend direction
  - Notable peaks and drops
  - Comparison between months
  - Forecasted future performance
- **Interactivity:** Updates dynamically based on filters applied on the page

Page 4: Sales Key Influencers

Key Influencers Visual

Title

Sales Key Influencers

Purpose

Helps identify and understand the factors that most strongly impact Sales. This visual explains *why* sales increase or decrease by analyzing relationships between Sales and other fields.

Visualization Details:

- **Analyze:** Sales

- **Explain By:**
  - Sum of Quantity
  - Price
  - Product Line

What the Key Influencers Visual Shows

- **Influential Factors:**

  The visual ranks which factors have the biggest effect on Sales.

  For example, it may show that higher quantities purchased or certain product lines are strongly linked to higher sales.

- **Driver Strength:**

  Each influencer displays how much it increases or decreases Sales.

  Power BI uses statistical analysis to determine which fields have the strongest relationship.

- **Top Segments View:**

  Shows groups of customers or products that share similar characteristics and have unusually high or low sales.

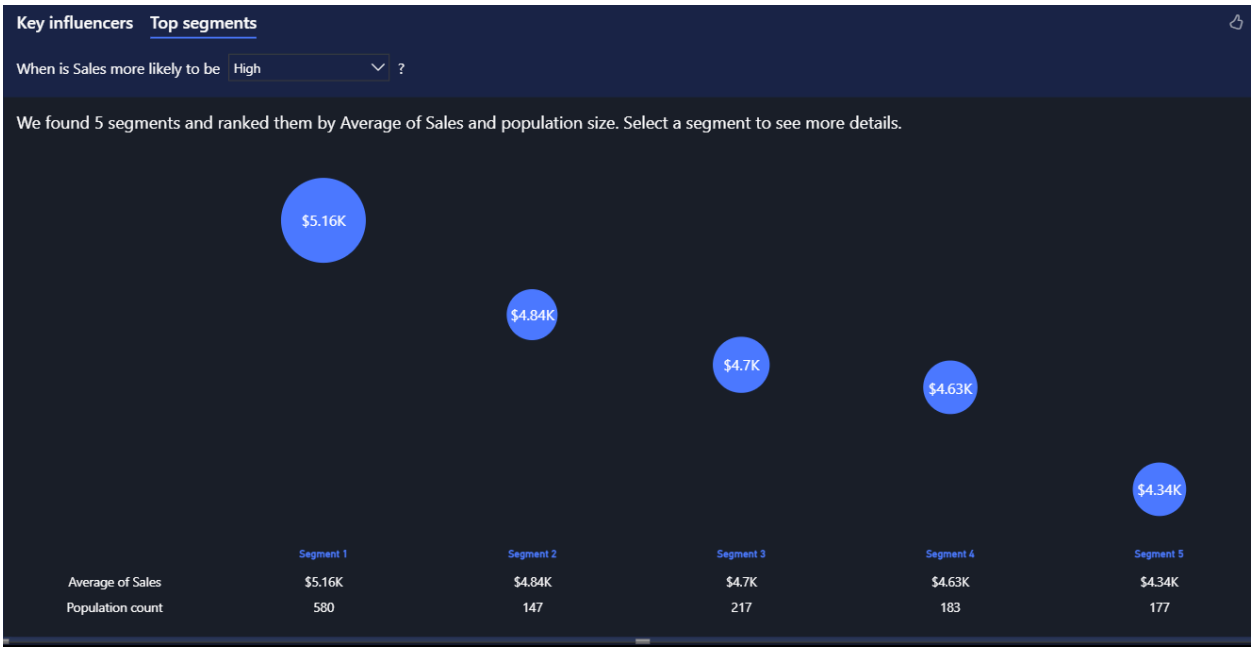
  This helps reveal hidden patterns—such as a specific product line performing better among certain customer groups.

- **Visualization Interpretation:**
  - Bars indicate how strongly each factor influences Sales.
  - Larger bars = stronger impact.
  - You can click influencers to see detailed breakdowns.

- **Dynamic Behavior:**

  The analysis updates automatically based on filters or selections applied in the report.


Key influencers
Top segments
What influences Sales to
Increase
?
When..
...the average of Sales increases by
Sum of Each Quantity Price goes up 20.17
$1.17K
Product Line is Classic Cars
$759.7
On average when Sum of Each Quantity Price increases, Sales also increases.
Average of Sales
$7K
$6K
$5K
$4K
$3K
$2K
$1K
$0K
20
40
60
80
100
Each Quantity Price


Key influencers
Top segments
When is Sales more likely to be
High
?
We found 5 segments and ranked them by Average of Sales and population size. Select a segment to see more details.
$5.16K
$4.84K
$4.7K
$4.63K
$4.34K
Segment 1
Segment 2
Segment 3
Segment 4
Segment 5
Average of Sales
$5.16K
$4.84K
$4.7K
$4.63K
$4.34K
Population count
580
147
217
183
177

Page 6: Sales Exploration

Decomposition Tree

Title

Sales Decomposition Analysis

Purpose

Breaks down total Sales into contributing factors, allowing step-by-step exploration of what drives increases or decreases in Sales.

Visualization Details:

- **Analyze:** Sum of Sales
- **Explain By:**
  - Deal Size
  - Territory
  - Product Line
  - Customer Name
- **Functionality:**
  - Users expand branches to explore how each dimension contributes to total Sales
  - Shows highest/lowest contributor options
  - Helps identify patterns and root causes behind sales performance
- **Interactivity:**
  - Responds dynamically to slicer selections
  - Enables drill-down exploration in any order chosen by the user

Slicers

Year Slicer

Title

Year Filter

Purpose

Filters the entire page based on the selected year.

Visualization Details:

- **Data:** Year
- **Interactivity:** Affects the Decomposition Tree and all other visuals on the page

Month Slicer

Title

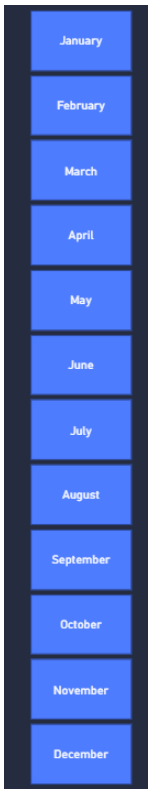
Month Filter

Purpose

Filters the page to show results for specific months within the selected year.

Visualization Details:

- **Data:** Month
- **Interactivity:**
  - Works together with the Year slicer
  - Updates the Decomposition Tree and all page visuals accordingly



Page 7: Next Year Sales Performance

Line Chart

Title

Next Year Sales Forecast

Purpose

Displays the predicted sales for the next 12 months using an Exponential Smoothing time-series forecasting model.

Visualization Details:

- **Data Source:** Forecast CSV file generated by a Python script (`monthly_sales_forecast1.csv`)
- **Columns Used:**
  - **ORDERDATE** – Future month-end dates
  - **SALES** – Forecasted sales values
- **Visualization Type:** Line Chart showing only forecasted monthly values
- **Axis:**
  - **X-Axis:** Order Date (future months only)
  - **Y-Axis:** Forecasted Sales
- **Data Scope:**
  - **Only future predictions are displayed**
  - Historical sales are **not** shown on this page
- **Interactivity:**
  - Responds to page-level filters if any exist

Python Forecasting Details

Overview

The forecast shown on this page is produced using a classical time-series forecasting model (**Holt-Winters Exponential Smoothing**) executed in Python.

How the Forecast Is Generated:

1. **Aggregate Sales by Month**

   Sales are grouped at a monthly level using the `ORDERDATE` column.

2. **Apply Exponential Smoothing Model**

   The script trains a Holt-Winters model with:

   - **Additive Trend**
   - **Additive Seasonality**
   - **12-month Seasonal Period**

3. **Forecast the Next 12 Months**

The model generates 12 future monthly predictions after the last available month.

4. **Create Forecast Dataset**

   The forecasted values are placed into a dataframe with:

   - Future dates (one per month)
   - Predicted sales values

5. **Export to CSV**

   The final file (`monthly_sales_forecast1.csv`) is loaded into Power BI and used to build the forecast line chart.

Python Code:

```
from statsmodels.tsa.holtwinters import ExponentialSmoothing
import pandas as pd
data = pd.read_csv('sales_data_sample.csv', encoding='latin1')

data['ORDERDATE'] = pd.to_datetime(data['ORDERDATE'], format='%m/%d/%
Y %H:%M')
monthly_sales = data.groupby(pd.Grouper(key='ORDERDATE', freq='M'))['SAL
ES'].sum()
model = ExponentialSmoothing(
    monthly_sales,
    trend='add',
    seasonal='add',
    seasonal_periods=12
).fit()

forecast = model.forecast(12)
future_dates = pd.date_range(
    start=monthly_sales.index[-1] + pd.DateOffset(months=1),
    periods=12,
    freq='M'
)
```

```
forecast_df = pd.DataFrame({
    'ORDERDATE': future_dates,
    'SALES': forecast.values
})
forecast_df.to_csv('monthly_sales_forecast2.csv', index=False)
```

Written by:

Nada Mohammad Kamal Alnajjar (202312960)