

## ASSIGNMENT 2:-

### Task 1: Understanding Power Query Editor (Data Cleaning & Transformation)

1. Import the dataset into Power BI using **Power Query Editor**.

2. Perform the following transformations:

- I. Remove missing values & duplicates.
- II. Change data types (e.g., Date, Currency, Categories).
- III. Split full names into first and last names.
- IV. Merge tables based on keys (e.g., Orders & Customers).

3. Save a **step-by-step documentation** of transformations done.



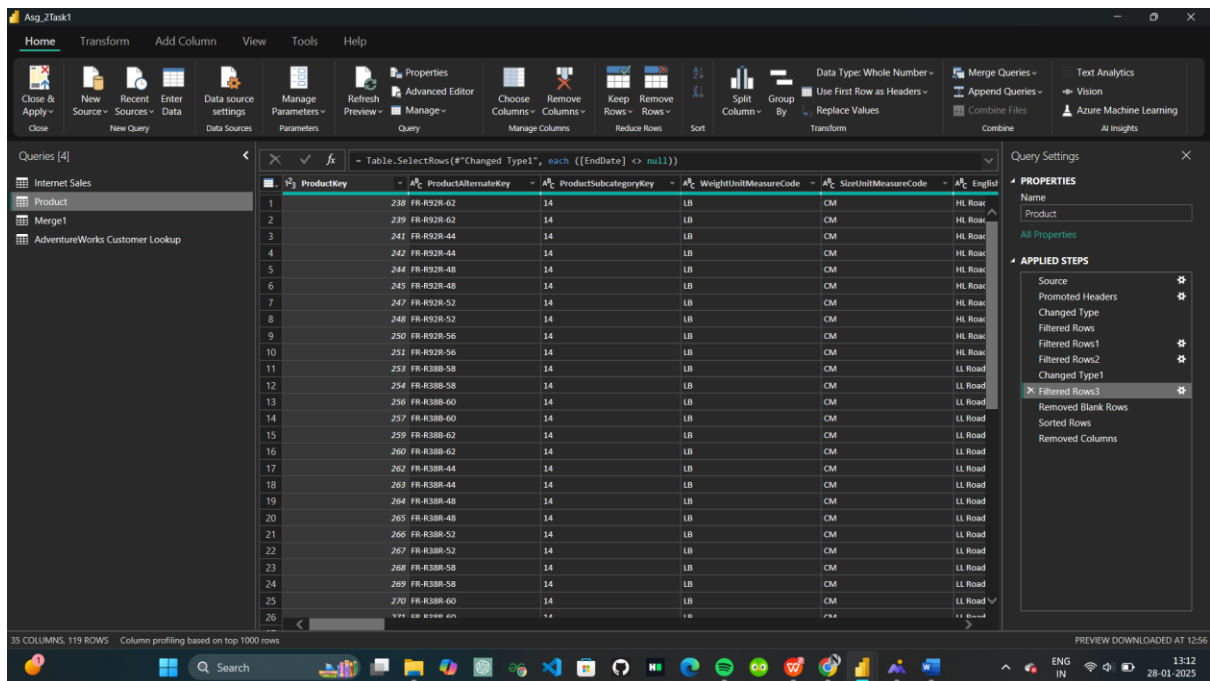
- Removed duplicates, NULL values, missing values :- On Product Dataset
- Sorted the Product key in ascending order

Before:

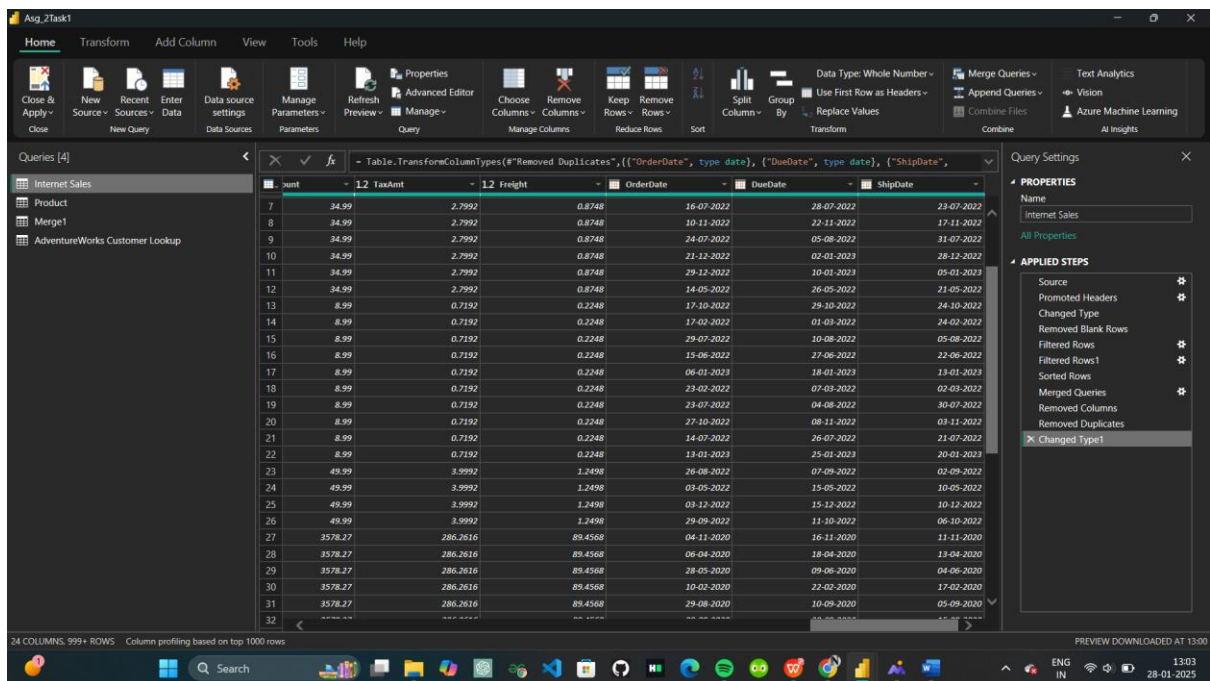
The screenshot displays the Power Query Editor window. The main area shows a table with 26 rows and 6 columns. The columns are: ProductKey, ProductAlternateKey, ProductSubcategoryKey, WeightUnitMeasureCode, SizeUnitMeasureCode, and English. The data is sorted by ProductKey in ascending order. The right sidebar shows the 'APPLIED STEPS' list, which includes: Source, Promoted Headers, Changed Type, Filtered Rows, Filtered Rows1, Filtered Rows2, Changed Type1, Filtered Rows3, Removed Blank Rows, Sorted Rows, and Removed Columns. The bottom status bar indicates '35 COLUMNS, 233 ROWS' and 'Column profiling based on top 1000 rows'.

	ProductKey	ProductAlternateKey	ProductSubcategoryKey	WeightUnitMeasureCode	SizeUnitMeasureCode	English
1	238	FR-R328-62	14	LB	CM	HL Road
2	239	FR-R328-62	14	LB	CM	HL Road
3	240	FR-R328-62	14	LB	CM	HL Road
4	241	FR-R328-44	14	LB	CM	HL Road
5	242	FR-R328-44	14	LB	CM	HL Road
6	243	FR-R328-44	14	LB	CM	HL Road
7	244	FR-R328-48	14	LB	CM	HL Road
8	245	FR-R328-48	14	LB	CM	HL Road
9	246	FR-R328-48	14	LB	CM	HL Road
10	247	FR-R328-52	14	LB	CM	HL Road
11	248	FR-R328-52	14	LB	CM	HL Road
12	249	FR-R328-52	14	LB	CM	HL Road
13	250	FR-R328-56	14	LB	CM	HL Road
14	251	FR-R328-56	14	LB	CM	HL Road
15	252	FR-R328-56	14	LB	CM	HL Road
16	253	FR-R388-58	14	LB	CM	LL Road
17	254	FR-R388-58	14	LB	CM	LL Road
18	255	FR-R388-58	14	LB	CM	LL Road
19	256	FR-R388-60	14	LB	CM	LL Road
20	257	FR-R388-60	14	LB	CM	LL Road
21	258	FR-R388-60	14	LB	CM	LL Road
22	259	FR-R388-62	14	LB	CM	LL Road
23	260	FR-R388-62	14	LB	CM	LL Road
24	261	FR-R388-62	14	LB	CM	LL Road
25	262	FR-R388-44	14	LB	CM	LL Road
26	263	FR-R388-44	14	LB	CM	LL Road

After:



- Changed the datatypes for Orderdate DueDate and ShipDate as date:-



- Split full names into first and last names. On Adventure Works Customer Lookup Dataset

**Split Column by Delimiter**

Specify the delimiter used to split the text column.

Select or enter delimiter  
Comma

Split at  
☐ Left-most delimiter  
☐ Right-most delimiter  
☒ Each occurrence of the delimiter

Advanced options  
 Quote Character  
 None

☐ Split using special characters  
 Insert special character:

OK Cancel

CustomerKey	Prefix	FirstName	LastName	BirthDate	MaritalStatus
11000	MRL	JON	YANG	08-04-1960	M
11001	MRL	EUGENE	HUANG	14-05-1965	S
11002	MRL	RUBEN	TORRES	12-08-1965	M
11003	MSL	CHRISTY	ZHU	15-02-1968	S
11004	MRS.	ELIZABETH	JOHNSON	08-08-1968	S
11005	MRL	JULIO	RUIZ	05-08-1965	S
11007	MRL	MARCO	MEHTA	09-05-1964	M
11008	MRS.	ROBIN	VERHOFF	07-07-1964	S
11009	MRL	SHANNON	CARLSON	01-04-1964	S
11010	MSL	JACQUELYN	SUAREZ	06-02-1964	S
11011	MRL	CURTIS	LU	04-11-1963	M
11012	MRS.	LAUREN	WALKER	18-01-1968	M
11013	MRL	IAN	JENKINS	06-08-1968	M
11014	MRS.	SYDNEY	BENNETT	09-05-1968	S
11015	MSL	CHLOE	YOUNG	27-02-1979	S
11016	MRL	WYATT	HILL	28-04-1979	M
11017	MRS.	SHANNON	WANG	26-06-1944	S
11018	MRL	CLARENCE	RAI	09-10-1944	S
11019	MRL	LUKE	LAL	07-03-1978	S
11020	MRL	JORDAN	KING	20-09-1978	S
11021	MSL	DESTINY	WILSON	03-09-1978	S
11022	MRL	ETHAN	ZHANG	12-10-1978	M
11023	MRL	SETH	EDWARDS	11-10-1978	M
11024	MRL	RUSSELL	XIE	17-09-1978	M
11025	MRL	ALEJANDRO	BECK	23-12-1945	M

**Table.RenameColumns**(#"Changed Type1",{"Fullname.1", "FirstName"}, {"Fullname.2", "LastName"})

CustomerKey	Prefix	FirstName	LastName	BirthDate	MaritalStatus
11000	MRL	JON	YANG	08-04-1960	M
11001	MRL	EUGENE	HUANG	14-05-1965	S
11002	MRL	RUBEN	TORRES	12-08-1965	M
11003	MSL	CHRISTY	ZHU	15-02-1968	S
11004	MRS.	ELIZABETH	JOHNSON	08-08-1968	S
11005	MRL	JULIO	RUIZ	05-08-1965	S
11007	MRL	MARCO	MEHTA	09-05-1964	M
11008	MRS.	ROBIN	VERHOFF	07-07-1964	S
11009	MRL	SHANNON	CARLSON	01-04-1964	S
11010	MSL	JACQUELYN	SUAREZ	06-02-1964	S
11011	MRL	CURTIS	LU	04-11-1963	M
11012	MRS.	LAUREN	WALKER	18-01-1968	M
11013	MRL	IAN	JENKINS	06-08-1968	M
11014	MRS.	SYDNEY	BENNETT	09-05-1968	S
11015	MSL	CHLOE	YOUNG	27-02-1979	S
11016	MRL	WYATT	HILL	28-04-1979	M
11017	MRS.	SHANNON	WANG	26-06-1944	S
11018	MRL	CLARENCE	RAI	09-10-1944	S
11019	MRL	LUKE	LAL	07-03-1978	S
11020	MRL	JORDAN	KING	20-09-1978	S
11021	MSL	DESTINY	WILSON	03-09-1978	S
11022	MRL	ETHAN	ZHANG	12-10-1978	M
11023	MRL	SETH	EDWARDS	11-10-1978	M
11024	MRL	RUSSELL	XIE	17-09-1978	M
11025	MRL	ALEJANDRO	BECK	23-12-1945	M

- Merged internet sales and products:- Taking Product key as a common key from both the tables

The screenshot shows the 'Merge' dialog box in Power Query. The 'Internet Sales' table is selected as the first table, and the 'Product' table is selected as the second table. The 'ProductKey' column is chosen as the common key for both tables. The 'Join Kind' is set to 'Left Outer (all from first, matching from second)'. The 'Fuzzy matching' options are checked. The 'The selection matches 726 of 1123 rows from the first table.' message is displayed. The 'OK' button is highlighted.

The screenshot shows the 'Merge' dialog box in Power Query. The 'Internet Sales' table is selected as the first table, and the 'Product' table is selected as the second table. The 'ProductKey' column is chosen as the common key for both tables. The 'Join Kind' is set to 'Left Outer (all from first, matching from second)'. The 'Fuzzy matching' options are checked. The 'The selection matches 726 of 1123 rows from the first table.' message is displayed. The 'OK' button is highlighted.

## Task 2: Understanding Data & Data Modeling

1. Identify key tables: Fact (Sales, Orders) and Dimension (Customers, Products, Regions).
2. Define **Primary & Foreign Keys** and create relationships in **Model View**.
3. Implement **Star Schema** for better performance.
4. Explain why schema design is important in a short document (1-2 pages).

### Fact Tables :

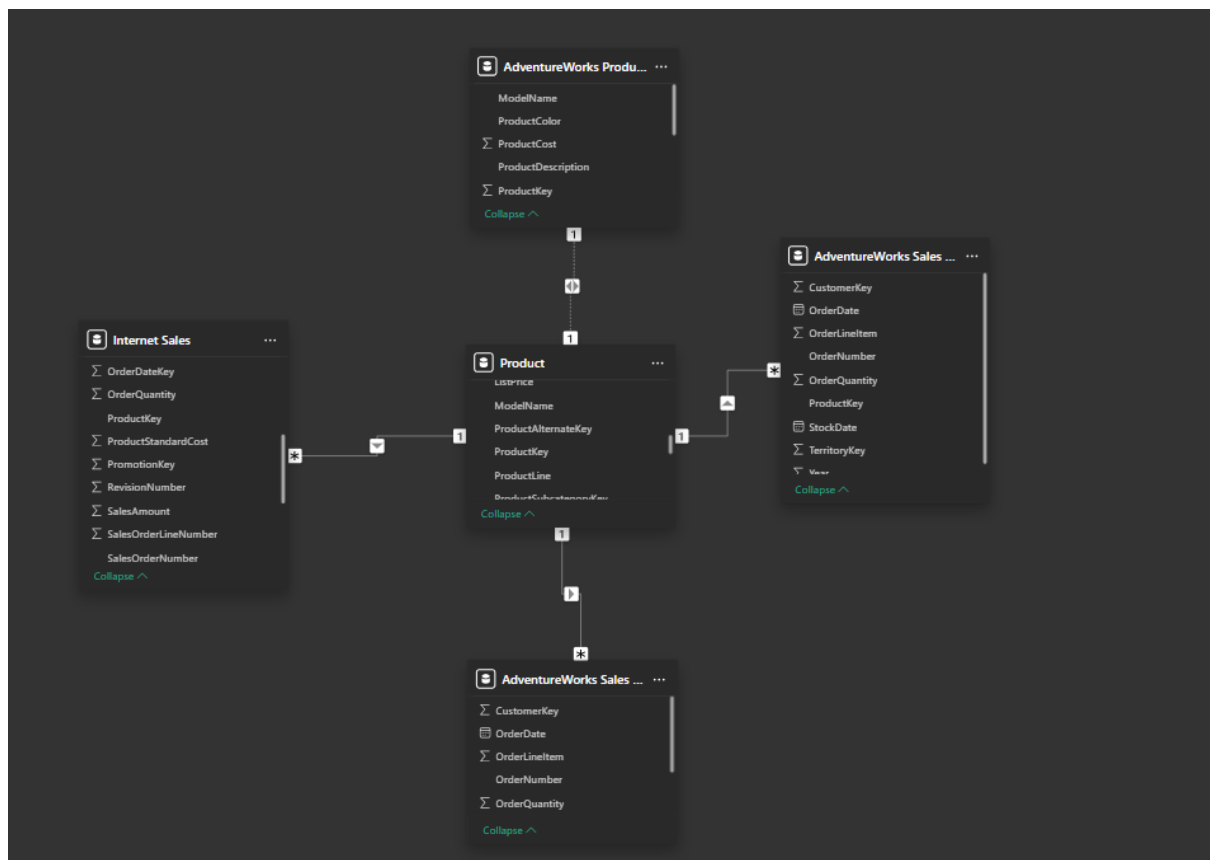
A fact table is the central table in a star schema

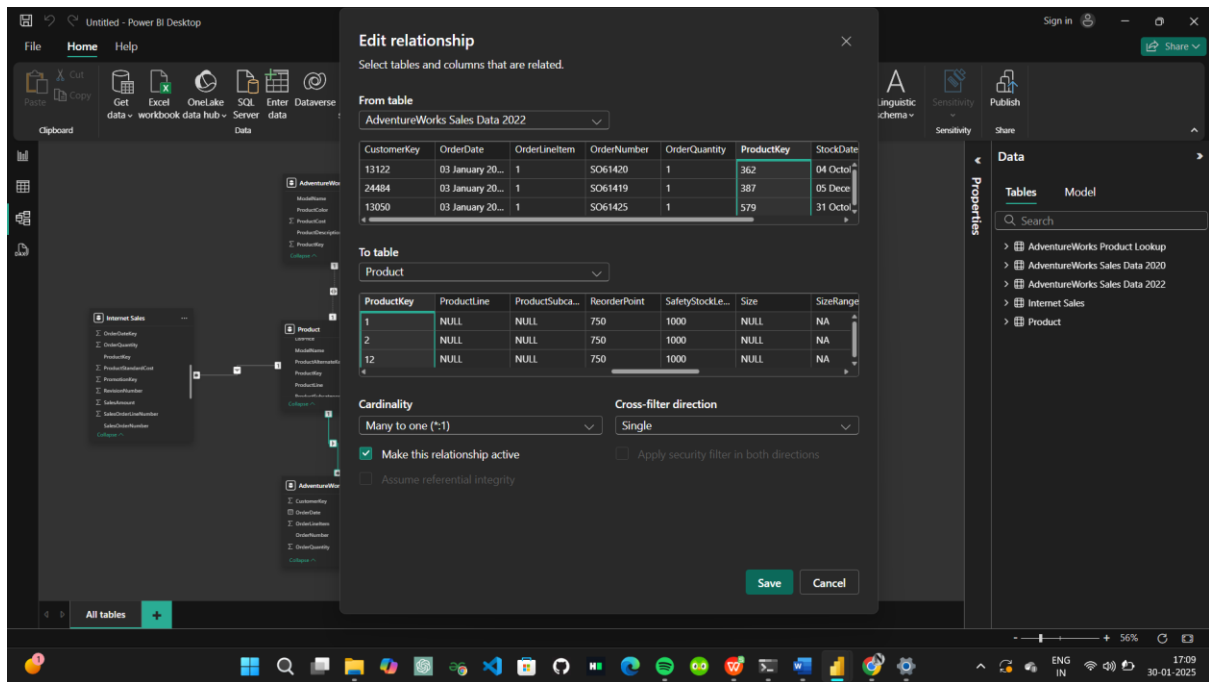
- Product

### Dimension Tables:

A dimension table provides the contextual information about the data stored in the fact table. It contains descriptive attributes (also called dimensions)

- AdventureWorks Product lookup
- AdventureWorks Sales Data 2020
- AdventureWorks Sales Data 2020
- Internet sales





- Relationships: Many-to-One

### Task 3: Implementing DAX Functions

#### A) Text Functions Create calculated columns using the following **Text DAX Functions**:

1. Extract the first 3 letters of Product Category:
2. Create a full name column from First & Last Names:

#### B) Logical Functions

Use **Logical DAX Functions** for analysis:

1. **Discount Category**: Create a calculated column to categorize discounts
2. High-Value Customers:

#### C) Time & Intelligence Functions

Use **Time Intelligence DAX Functions** to analyze sales trends:

1. Calculate **Total Sales Last Year**:
2. Find **Year-to-Date (YTD) Sales**:
3. Calculate **Month-over-Month Growth**



#### A) Text Functions Create calculated columns using the following **Text DAX Functions**:

1. Extract the first 3 letters of Product Category: (DataSet:-Product Category)

First3Letters\_English = LEFT('Product Category'[EnglishProductCategoryName],3)

First3Letters\_Spanish = LEFT('Product Category'[SpanishProductCategoryName],3)

First3Letters\_French = LEFT('Product Category'[FrenchProductCategoryName],3)

The screenshot shows the Microsoft Power BI Desktop interface. The 'Column tools' ribbon is active, displaying the formula bar with the DAX formula: `First3Letters_English = LEFT('Product Category'[EnglishProductCategoryName],3)`. Below the formula bar, a data table is visible with the following columns: ProductCategoryAlternateKey, EnglishProductCategoryName, SpanishProductCategoryName, FrenchProductCategoryName, First3Letters\_English, First3Letters\_Spanish, and First3Letters\_French. The table contains 4 rows of data. The right-hand pane shows the 'Data' view with a search bar and a list of fields, including the newly created calculated columns.

ProductCategoryAlternateKey	EnglishProductCategoryName	SpanishProductCategoryName	FrenchProductCategoryName	First3Letters_English	First3Letters_Spanish	First3Letters_French
1	Bikes	Bicicleta	Vélo	Blk	Bic	Vél
2	Components	Componente	Composant	Com	Com	Com
3	Clothing	Prenda	Vêtements	Clo	Pre	Vêt
4	Accessories	Accesorio	Accessoire	Acc	Acc	Acc

## 2. Create a full name column from First & Last Names:

FullName = COMBINEVALUES(" ", 'AdventureWorks Customer Lookup'[FirstName], 'AdventureWorks Customer Lookup'[LastName])

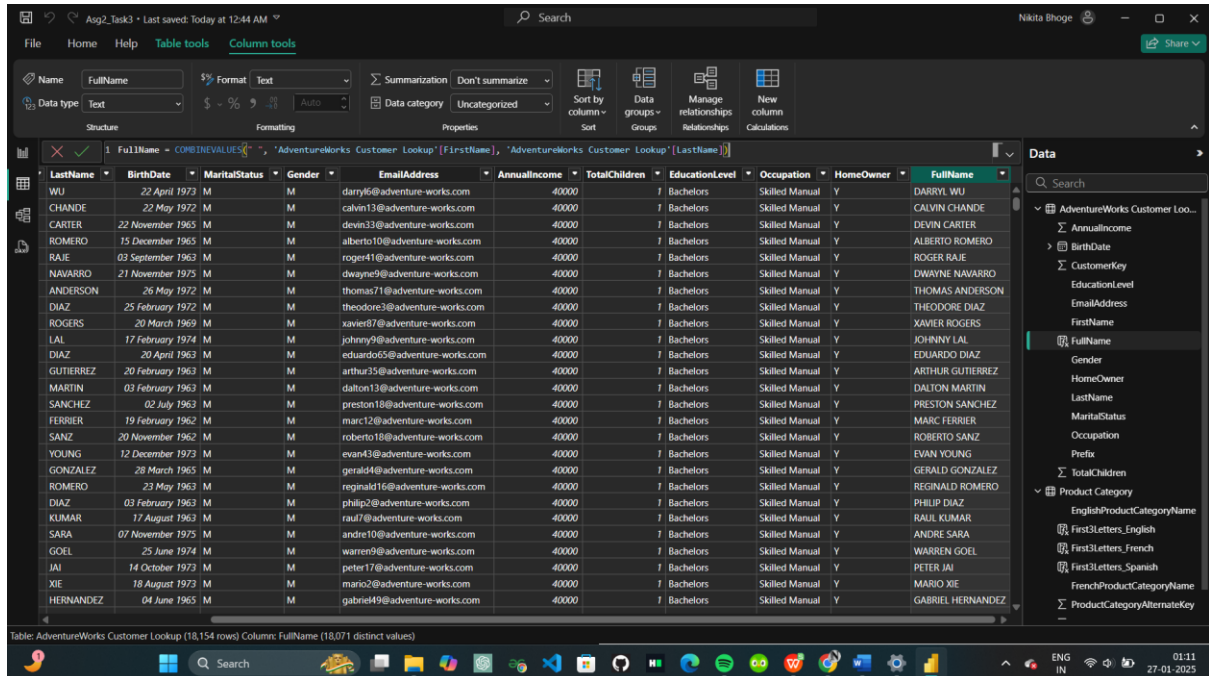


Table: AdventureWorks Customer Lookup (18,154 rows) Column: FullName (18,071 distinct values)

LastName	BirthDate	MaritalStatus	Gender	EmailAddress	AnnualIncome	TotalChildren	EducationLevel	Occupation	HomeOwner	FullName
WU	22 April 1973	M	M	darryl6@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	DARRYL WU
CHANDE	22 May 1972	M	M	calvin13@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	CALVIN CHANDE
CARTER	22 November 1965	M	M	devin33@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	DEVIN CARTER
ROMERO	15 December 1965	M	M	alberto10@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	ALBERTO ROMERO
RAJE	03 September 1963	M	M	roger41@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	ROGER RAJE
NAVARRO	21 November 1975	M	M	dwayne9@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	DWAYNE NAVARRO
ANDERSON	26 May 1972	M	M	thomas71@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	THOMAS ANDERSON
DIAZ	25 February 1972	M	M	theodore3@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	THEODORE DIAZ
ROGERS	20 March 1969	M	M	xavier87@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	XAVIER ROGERS
LAL	17 February 1974	M	M	johnny9@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	JOHNNY LAL
DIAZ	20 April 1963	M	M	eduardo65@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	EDUARDO DIAZ
GUTIERREZ	20 February 1963	M	M	arthur35@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	ARTHUR GUTIERREZ
MARTIN	03 February 1963	M	M	dalton13@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	DALTON MARTIN
SANCHEZ	02 July 1963	M	M	preston18@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	PRESTON SANCHEZ
FERRIER	19 February 1962	M	M	marc12@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	MARC FERRIER
SANZ	20 November 1962	M	M	roberto18@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	ROBERTO SANZ
YOUNG	12 December 1973	M	M	evan43@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	EVAN YOUNG
GONZALEZ	28 March 1965	M	M	gerald4@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	GERALD GONZALEZ
ROMERO	23 May 1963	M	M	reginald16@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	REGINALD ROMERO
DIAZ	03 February 1963	M	M	philip2@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	PHILIP DIAZ
KUMAR	17 August 1963	M	M	raul7@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	RAUL KUMAR
SARA	07 November 1975	M	M	andre10@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	ANDRE SARA
GOEL	25 June 1974	M	M	warren9@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	WARREN GOEL
JAI	14 October 1973	M	M	peter17@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	PETER JAI
XIE	18 August 1973	M	M	mario2@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	MARIO XIE
HERNANDEZ	04 June 1965	M	M	gabriel49@adventure-works.com	40000	7	Bachelors	Skilled Manual	Y	GABRIEL HERNANDEZ



## B) Logical Functions

Use **Logical DAX Functions** for analysis:

1. **Discount Category:** Create a calculated column to categorize discounts

DiscountedPrice = 'AdventureWorks Product Lookup'[ProductPrice]\*0.75

DiscountAmount = 'AdventureWorks Product Lookup'[ProductPrice]\*0.25

DiscountCategory = IF('AdventureWorks Product Lookup'[DiscountAmount]<50,"Low",IF('AdventureWorks Product Lookup'[DiscountAmount]>50,"High"))

Profit = (SUM('AdventureWorks Product Lookup'[DiscountedPrice])-SUM('AdventureWorks Product Lookup'[ProductCost]))

TotalSales = SUM('AdventureWorks Product Lookup'[DiscountedPrice])

Table: AdventureWorks Product Lookup (293 rows) Column: DiscountedPrice (114 distinct values)

ProductColor	ProductSize	ProductStyle	ProductCost	ProductPrice	DiscountedPrice	DiscountAmount	DiscountCategory	CustomerCategory	TotalSales	Profit
Red	0	0	13.0863	55.99	41.9925	13.9975	Low	Low Value	180481.609125	28.9062
Black	0	0	200.0278	563.6442	422.73315	140.91105	High	High Value	180481.609125	222.70535
Blue	0	0	112.0278	784.6442	588.48315	196.16105	High	High Value	180481.609125	476.45535
Multi	0	U	234.7052	484.6442	363.48315	121.16105	High	High Value	180481.609125	128.77795
NA	0	0	65.8097	148.22	111.165	37.055	Low	High Value	180481.609125	45.3553
NA	0	0	77.9176	175.49	131.6175	43.8725	Low	High Value	180481.609125	53.6999
NA	0	0	101.8936	229.49	172.1175	57.3725	High	High Value	180481.609125	70.2239
NA	0	0	15.1848	34.2	25.65	8.55	Low	Low Value	180481.609125	10.4852
NA	0	0	45.4168	102.29	76.7175	25.5725	Low	High Value	180481.609125	31.3007
NA	0	0	55.3801	124.73	93.5475	31.1825	Low	High Value	180481.609125	38.1674
NA	0	0	17.978	40.4909	30.368175	10.122725	Low	Low Value	180481.609125	12.390175
NA	0	0	24.9932	56.2909	42.218175	14.077225	Low	Low Value	180481.609125	17.224975
NA	0	0	48.5453	109.3364	82.0023	27.3341	Low	High Value	180481.609125	33.457
NA	0	0	17.978	40.4909	30.368175	10.122725	Low	Low Value	180481.609125	12.390175
NA	0	0	24.9932	56.2909	42.218175	14.077225	Low	Low Value	180481.609125	17.224975
NA	0	0	48.5453	109.3364	82.0023	27.3341	Low	High Value	180481.609125	33.457
Black	0	0	26.9708	60.745	45.55875	15.18625	Low	Low Value	180481.609125	18.58795
Black	0	0	92.8071	209.025	156.76875	52.25625	High	High Value	180481.609125	63.96165
Black	0	0	133.2955	300.215	222.16125	75.05375	High	High Value	180481.609125	91.86575
Black	0	0	37.9909	85.565	64.17375	21.39125	Low	Low Value	180481.609125	26.18285
Black	0	0	110.2829	248.385	186.28875	62.09625	High	High Value	180481.609125	76.00585
Black	0	0	146.5466	330.06	247.545	82.515	High	High Value	180481.609125	100.9984
Black	0	0	96.7964	218.01	163.5075	54.5025	High	High Value	180481.609125	66.7111
Black	0	0	38.9588	87.745	65.80875	21.93625	Low	Low Value	180481.609125	26.84995
Black	0	0	104.7951	236.025	177.01875	59.00625	High	High Value	180481.609125	72.22365
Black	0	0	145.2835	327.215	245.41125	81.80375	High	High Value	180481.609125	100.12775
Black	0	0	49.9789	112.565	84.42375	28.14125	Low	High Value	180481.609125	34.44485

Asp2\_Task3 • Last saved: Today at 1:40 AM

Search

File Home Help Table tools Column tools

Name DiscountCategory Format Text

Data type Text

Structure

Summarization Don't summarize

Data category Uncategorized

Sort by column

Sort

Data groups

Manage relationships

New column

Calculations

DiscountCategory = IF('AdventureWorks Product Lookup'[DiscountAmount]<50,"Low",IF('AdventureWorks Product Lookup'[DiscountAmount]>50,"High"))

ProductDescription	ProductColor	ProductSize	ProductStyle	ProductCost	ProductPrice	DiscountedPrice	DiscountAmount	DiscountCategory
visor.	Red	0	0	73.0863	55.99	41.9925	13.9975	Low
visor.	Black	0	0	200.0278	563.6442	422.73315	140.91105	High
visor.	Blue	0	0	112.0278	784.6442	588.48315	196.16105	High
s all.	Multi	0	U	234.7052	484.6442	363.48315	121.16105	High
ecise steering.	NA	0	0	65.8097	148.22	111.165	37.055	Low
tube.	NA	0	0	77.9176	175.49	131.6175	43.8725	Low
d legs.	NA	0	0	101.8936	229.49	172.1175	57.3725	High
omical price.	NA	0	0	15.1848	34.2	25.65	8.55	Low
	NA	0	0	45.4168	102.29	76.7175	25.5725	Low
e port for quick lubrication.	NA	0	0	55.3801	124.73	93.5475	31.1825	Low
	NA	0	0	17.978	40.4909	30.368175	10.122725	Low
	NA	0	0	24.9932	56.2909	42.218175	14.072725	Low
	NA	0	0	48.5453	109.3364	82.0023	27.3341	Low
ts.	NA	0	0	17.978	40.4909	30.368175	10.122725	Low
it all riders.	NA	0	0	24.9932	56.2909	42.218175	14.072725	Low
ped bar from aluminum alloy.	NA	0	0	48.5453	109.3364	82.0023	27.3341	Low
ter.	Black	0	0	26.9708	60.745	45.55875	15.18625	Low
serious rider.	Black	0	0	92.8071	209.025	156.76875	52.25625	High
tl.	Black	0	0	133.2955	300.215	225.16125	75.05375	High
yclist.	Black	0	0	37.9909	65.585	64.17375	21.39125	Low
	Black	0	0	110.2829	248.385	186.28875	62.09625	High
	Black	0	0	186.5466	338.06	247.545	90.515	High
	Black	0	0	96.7964	218.01	163.5075	54.5025	High
ter.	Black	0	0	38.8508	87.745	65.98875	21.93625	Low
serious rider.	Black	0	0	104.7951	236.025	177.01875	59.00625	High
sl.	Black	0	0	145.2835	327.215	245.41125	81.80375	High

Table: AdventureWorks Product Lookup (293 rows) Column: DiscountCategory (3 distinct values)

Search

MaritalStatus

Occupation

Prefix

TotalChildren

AdventureWorks Product Lookup

DiscountAmount

DiscountCategory

DiscountedPrice

ModelName

ProductColor

ProductCost

ProductKey

ProductDescription

ProductName

ProductPrice

ProductSize

ProductSKU

ProductStyle

ProductSubcategoryKey

Internet Sales

CarrierTrackingNumber

CurrencyKey

ENG IN

01:41

27-01-2025

## 2. High-Value Customers:

CustomerCategory = IF('AdventureWorks Product Lookup'[DiscountedPrice]>70,"High Value",IF('AdventureWorks Product Lookup'[DiscountedPrice]<70,"Low Value"))

The screenshot shows an Excel spreadsheet with a DAX formula in the 'CustomerCategory' column. The formula is: `=IF('AdventureWorks Product Lookup'[DiscountedPrice]>70,"High Value",IF('AdventureWorks Product Lookup'[DiscountedPrice]<70,"Low Value"))`. The data table has columns: ProductColor, ProductSize, ProductStyle, ProductCost, ProductPrice, DiscountedPrice, DiscountAmount, DiscountCategory, and CustomerCategory. The data is filtered by 'High Value' customers.

ProductColor	ProductSize	ProductStyle	ProductCost	ProductPrice	DiscountedPrice	DiscountAmount	DiscountCategory	CustomerCategory
Red	0	0	13.0863	55.99	41.9925	13.9975	Low	Low Value
Black	0	0	200.0278	563.6442	422.73315	140.91105	High	High Value
Blue	0	0	112.0278	784.6442	588.48315	196.16105	High	High Value
Multi	0	U	234.7052	484.6442	363.48315	121.16105	High	High Value
NA	0	0	65.8097	148.22	111.165	37.055	Low	High Value
NA	0	0	72.9176	175.49	131.6175	43.8725	Low	High Value
NA	0	0	101.8936	228.49	172.1175	52.3725	High	High Value
NA	0	0	15.1848	34.2	25.65	8.55	Low	Low Value
NA	0	0	45.4168	102.29	76.7175	25.5725	Low	High Value
NA	0	0	55.3801	124.73	93.5475	31.1825	Low	High Value
NA	0	0	17.978	40.4909	30.368175	10.122725	Low	Low Value
NA	0	0	24.9932	56.2909	42.218175	14.072725	Low	Low Value
NA	0	0	48.5453	109.3364	82.0023	27.3341	Low	High Value
NA	0	0	17.978	40.4909	30.368175	10.122725	Low	Low Value
NA	0	0	24.9932	56.2909	42.218175	14.072725	Low	Low Value
NA	0	0	48.5453	109.3364	82.0023	27.3341	Low	High Value
Black	0	0	26.9708	60.745	45.5875	15.15825	Low	Low Value
Black	0	0	92.8071	209.025	156.76875	52.25625	High	High Value
Black	0	0	133.2955	300.215	225.16125	75.05375	High	High Value
Black	0	0	37.9909	85.565	64.17375	21.39225	Low	Low Value
Black	0	0	110.2829	248.385	186.28875	62.09625	High	High Value
Black	0	0	146.5466	330.06	247.545	82.515	High	High Value
Black	0	0	96.7984	218.01	163.5075	54.5025	High	High Value
Black	0	0	38.3580	87.745	65.80075	21.94425	Low	Low Value
Black	0	0	104.7951	236.025	177.01825	58.00625	High	High Value
Black	0	0	145.2835	327.215	245.41125	81.80375	High	High Value

## C) Time & Intelligence Functions

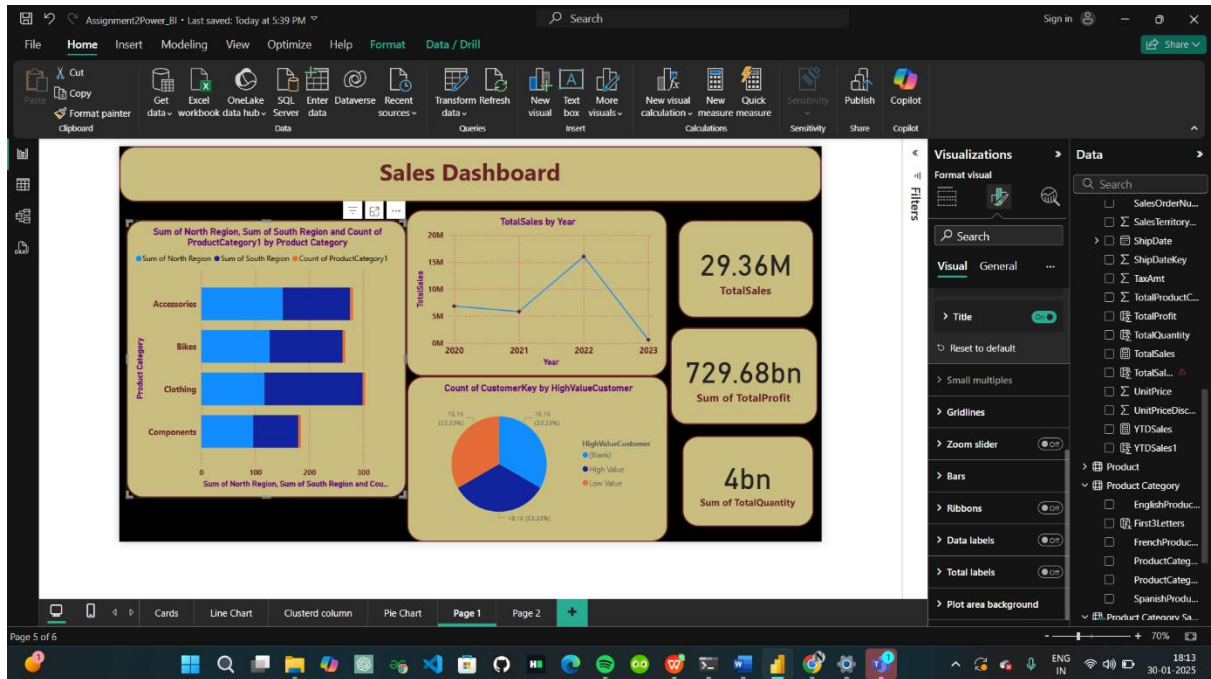
Use Time Intelligence DAX Functions to analyze sales trends:

1. Calculate Total Sales Last Year:
2. Find Year-to-Date (YTD) Sales:
3. Calculate Month-over-Month Growth

## Task 4: Data Visualization & Report Creation

Create a Sales Dashboard:

- Total Sales, Profit, and Quantity (Cards)
- Sales Trends over time (Line Chart)
- Sales by Region & Product Category (Bar Chart)
- High-Value vs Regular Customers (Pie Chart)



### Task 5: Submission Requirements

- Power BI Report (.pbix) File
- Step-by-step Documentation (for Power Query, Data Modeling, and DAX Functions)
- Summary Report (2-3 pages explaining insights & findings)

## Summary Report

### Task 1: Understanding Power Query Editor (Data Cleaning & Transformation)

In this task, the focus is on data cleaning and transformation using Power Query Editor:

1. **Data Import:** The dataset (AdventureWorks or E-commerce Sales) was imported into Power BI using Power Query Editor for further processing.
2. **Data Transformations:**
  - **Missing Values & Duplicates:** We removed any missing values and duplicates to ensure data integrity.
  - **Changing Data Types:** Proper data types (such as Date, Currency, and Categories) were assigned to columns for correct analysis.
  - **Splitting Full Names:** Full names were split into first and last names for easier analysis.
  - **Merging Tables:** Tables (e.g., Orders & Customers) were merged based on key fields, ensuring that related data could be analyzed together.
3. **Step-by-Step Documentation:** A detailed record of all transformations applied in Power Query Editor, along with any issues encountered and how they were resolved, was documented for future reference.

---

### Task 2: Understanding Data & Data Modeling

This task emphasizes data modeling and schema design:

1. **Identifying Key Tables:**
  - **Fact Tables:** Sales and Orders tables were identified as the Fact tables, containing transactional data.
  - **Dimension Tables:** Customers, Products, and Regions were classified as Dimension tables, providing descriptive data for analysis.
2. **Primary & Foreign Keys:** Relationships between tables were established by defining primary and foreign keys (e.g., CustomerID, ProductID) to link tables together in the data model.

3. **Star Schema:** The data model was designed using a Star Schema, where Fact tables are connected to multiple Dimension tables, simplifying the relationships and enhancing performance during queries.
  4. **Importance of Schema Design:**
    - The Star Schema was implemented to improve query performance by reducing complexity and enabling more efficient reporting. Proper schema design also ensures better data consistency and scalability.
- 

### Task 3: Implementing DAX Functions

DAX functions were applied to analyze and create calculated columns in the dataset:

#### A) Text Functions:

- **Extracting First 3 Letters of Product Category:** A calculated column was created to extract the first three letters of the Product Category using the LEFT DAX function.
- **Creating Full Name Column:** A new column was created by concatenating First and Last Names using DAX's CONCATENATE function.

#### B) Logical Functions:

- **Discount Category:** A calculated column was created to categorize discounts (e.g., High, Medium, Low) based on discount values using the IF and SWITCH functions.
- **High-Value Customers:** A logical function was used to identify high-value customers based on sales volume and profitability.

#### C) Time Intelligence Functions:

- **Total Sales Last Year:** The TOTALYTD function was used to calculate sales for the same period last year.
  - **Year-to-Date (YTD) Sales:** The DATESYTD function was used to calculate sales from the beginning of the year to the current date.
  - **Month-over-Month Growth:** DAX functions such as SAMEPERIODLASTYEAR and DIVIDE were used to calculate month-over-month sales growth.
- 

### Task 4: Data Visualization & Report Creation

This task focuses on creating interactive visual reports and dashboards:

#### 1. Sales Dashboard:

- **Total Sales, Profit, and Quantity:** Card visuals were created to display key metrics like Total Sales, Profit, and Quantity.
- **Sales Trends over Time:** A line chart was created to show how sales trends evolved over time, providing insights into sales performance across different periods.

- **Sales by Region & Product Category:** Bar charts were used to visualize sales data segmented by Region and Product Category, allowing for easy comparison.
- **High-Value vs Regular Customers:** A pie chart was created to compare the number of High-Value customers versus Regular customers.

## 2. Customer Analysis Dashboard:

- **Customer-wise Sales & Profits:** A table visual was created to show sales and profit details for each customer.
  - **Top 10 Customers:** A bar chart was created to highlight the top 10 customers based on sales.
  - **Discount Categories:** A stacked bar chart was created to show sales distribution across various discount categories.
3. **Filters & Slicers:** Dynamic filters and slicers were added to allow users to interact with the data, such as filtering by region, category, or time period.
-

