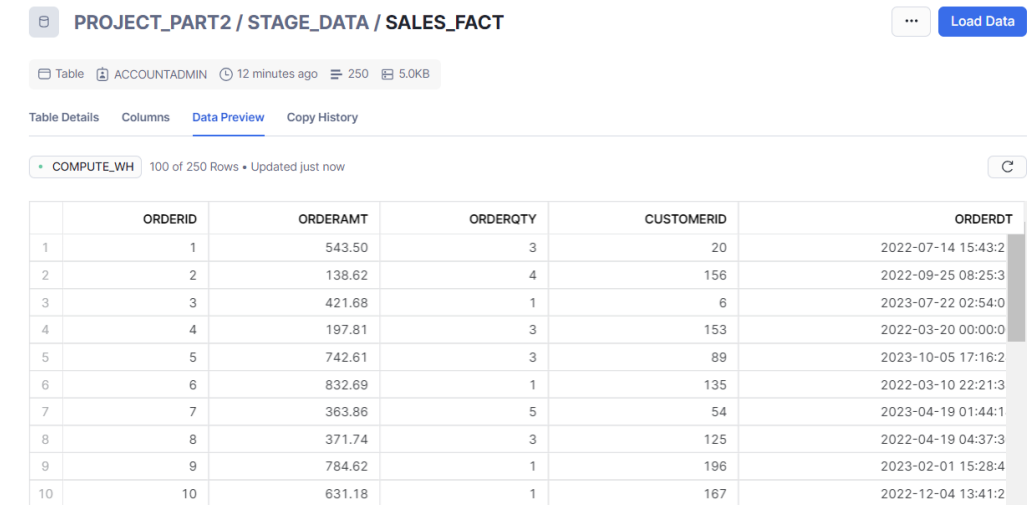
Project Part 2

Screenshot of SQL scripts of creating 3 Dimensional tables and 1 Fact Table in Stage Data.

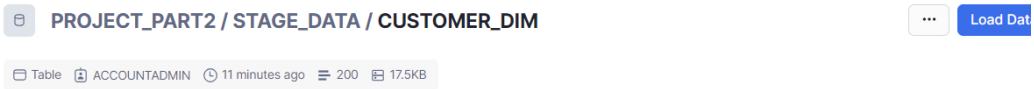
```
CREATE DATABASE PROJECT_PART2;

CREATE SCHEMA PROJECT_PART2.STAGE_DATA;


CREATE OR REPLACE TABLE Customer_Dim (
    CustomerID INT PRIMARY KEY,
    First_Name VARCHAR(255),
    Last_Name VARCHAR(255),
    Email VARCHAR(255),
    Phone VARCHAR(20),
    CustAddress VARCHAR(255),
    CustCity VARCHAR(100),
    CustState VARCHAR(50),
    CustZip VARCHAR(50)
);


CREATE OR REPLACE TABLE Product_Dim (
    ProductID INT PRIMARY KEY,
    Brand VARCHAR(200),
    Name VARCHAR(1000),
    Description VARCHAR(5000),
    Price DECIMAL(10, 2),
    MSRP DECIMAL(10, 2),
    Cost DECIMAL(10, 2)
);


CREATE OR REPLACE TABLE Sales_Fact (
    OrderID INT PRIMARY KEY,
    OrderAmt DECIMAL(12, 2),
```

```
    OrderQty INT,
    CustomerID INT,
    OrderDT TIMESTAMP,
    FOREIGN KEY (CustomerID) REFERENCES Customer_Dim(CustomerID)
);

CREATE OR REPLACE TABLE OrderDetail_Dim (
    OrderDetailID INT PRIMARY KEY,
    ProductID INT,
    OrderID INT,
    FOREIGN KEY (ProductID) REFERENCES Product_Dim(ProductID),
    FOREIGN KEY (OrderID) REFERENCES Sales_Fact(OrderID)
);

-- Partially data loaded for product dimension table, below are the remaining data added manually.

Insert into Product_Dim (
ProductID,Brand, Name, Description, Price, MSRP, Cost)
Values (1, 'TechWave', 'SpeedBoost Wireless Charging Pad', 'Fast efficient wireless power', 39.65, 44.41, 14.67),
(6, 'EcoEra', 'GreenGrow Indoor Herb Garden', 'Sustainable fresh home gardening', 209.99, 235.19,77.7
), (8, 'EcoEra', 'EarthGuard Plant-Based Cleaner', 'Natural powerful cleaning solution', 17.98, 20.14, 6.65
), (10, 'EcoEra', 'BioBlend Organic Protein Powder', 'Clean plant-powered nutrition boost', 89.99, 100.79, 33.3
), (12, 'LuxeLife', 'SilkSoothe Sleep Mask Set', 'Luxurious restful sleep experience', 45.99, 51.51, 17.02
);
```

Screenshot of data loading in Stage data:

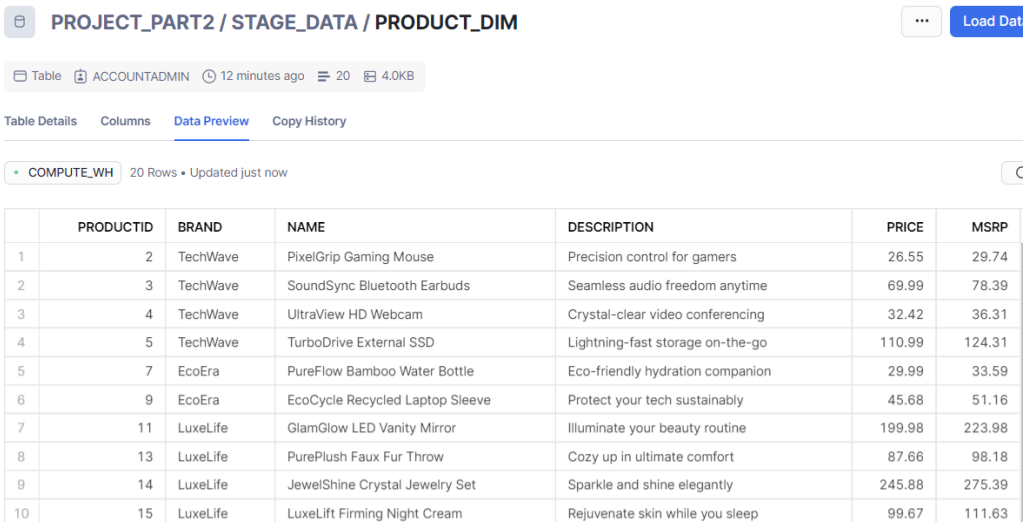


	ORDERID	ORDERAMT	ORDERQTY	CUSTOMERID	ORDERDT
1	1	543.50	3	20	2022-07-14 15:43:2
2	2	138.62	4	156	2022-09-25 08:25:3
3	3	421.68	1	6	2023-07-22 02:54:0
4	4	197.81	3	153	2022-03-20 00:00:0
5	5	742.61	3	89	2023-10-05 17:16:2
6	6	832.69	1	135	2022-03-10 22:21:3
7	7	363.86	5	54	2023-04-19 01:44:1
8	8	371.74	3	125	2022-04-19 04:37:3
9	9	784.62	1	196	2023-02-01 15:28:4
10	10	631.18	1	167	2022-12-04 13:41:2



	CUSTOMERID	FIRST_NAME	LAST_NAME	EMAIL	PHONE	CUSTADDRESS
1	1	Penelope	Keddie	pkeddie0@behance.net	205-151-7125	3 8th Crossing
2	2	Trip	Malyj	tmalyj1@oaic.gov.au	404-614-8569	13 Mcguire Road
3	3	Jake	Gorger	jgorger2@springer.com	208-562-9998	68153 Maryland Cen
4	4	Johnath	Game	jgame3@shareasale.com	410-276-1726	7994 Killdeer Parkwa
5	5	Angela	Nattriss	anattriss4@shinystat.com	513-792-4285	02 Maryland Hill
6	6	Linda	Jemmett	ljemmett5@slashdot.org	214-198-8022	6436 Gateway Alley
7	7	Lyman	Gascoyen	lgascoyen6@admin.ch	214-402-0630	045 Gale Street
8	8	Carlina	Shilliday	cshilliday7@shareasale.com	814-114-4016	708 Memorial Hill
9	9	Maribel	Guile	mguile8@techcrunch.com	808-103-4647	5421 Myrtle Lane
10	10	Della	Pisculli	dpisculli9@slashdot.org	530-256-6967	41 Jackson Pass

	ORDERDETAILID	PRODUCTID	ORDERID
1	1	213	16
2	2	60	6
3	3	188	12
4	4	229	16
5	5	121	12
6	6	115	13
7	7	39	3
8	8	127	1
9	9	115	7
10	10	159	20



	PRODUCTID	BRAND	NAME	DESCRIPTION	PRICE	MSRP
1	2	TechWave	PixelGrip Gaming Mouse	Precision control for gamers	26.55	29.74
2	3	TechWave	SoundSync Bluetooth Earbuds	Seamless audio freedom anytime	69.99	78.39
3	4	TechWave	UltraView HD Webcam	Crystal-clear video conferencing	32.42	36.31
4	5	TechWave	TurboDrive External SSD	Lightning-fast storage on-the-go	110.99	124.31
5	7	EcoEra	PureFlow Bamboo Water Bottle	Eco-friendly hydration companion	29.99	33.59
6	9	EcoEra	EcoCycle Recycled Laptop Sleeve	Protect your tech sustainably	45.68	51.16
7	11	LuxeLife	GlamGlow LED Vanity Mirror	Illuminate your beauty routine	199.98	223.98
8	13	LuxeLife	PurePlush Faux Fur Throw	Cozy up in ultimate comfort	87.66	98.18
9	14	LuxeLife	JewelShine Crystal Jewelry Set	Sparkle and shine elegantly	245.88	275.39
10	15	LuxeLife	LuxeLift Firming Night Cream	Rejuvenate skin while you sleep	99.67	111.63

Steps Taken to clean /Transform data.

1. Checked missing values for all the tables.
2. Dropped and filled missing values in Customer Dim Table.
3. Product, Order detail and Sales table have zero missing values.
4. Checked for Duplicate values to understand the difference in data shape of each table.
5. Updated Cust city column of Customer Dim table and removed inconsistency.
6. Validate the email format in the Email column of Customer_dim table.
7. Alter OrderDT in Sales Table to Year, Month, Day, Time, and AM/PM to interpret the data easily.
8. Changed the Month Numbers to Month Names in OrderDt_Month column under Sales Table.
9. Created calculated field “Total Sales” in Sales Table.
10. Created calculated field “Discount Rate” & “Markup Percentage” in Product_Dim Table.
11. Copied the Cleaned data from stage scheme to Prod Schema.

SQL Scripts for Data Cleaning:

```
-- Checking for Missing Values.
SELECT
    'Customer_Dim' AS Table_Name,
    SUM(CASE WHEN First_Name IS NULL THEN 1 ELSE 0 END) AS Missing_First_Name,
    SUM(CASE WHEN Last_Name IS NULL THEN 1 ELSE 0 END) AS Missing_Last_Name,
    SUM(CASE WHEN Email IS NULL THEN 1 ELSE 0 END) AS Missing_Email,
    SUM(CASE WHEN Phone IS NULL THEN 1 ELSE 0 END) AS Missing_Phone,
    SUM(CASE WHEN CustAddress IS NULL THEN 1 ELSE 0 END) AS Missing_CustAddress,
    SUM(CASE WHEN CustCity IS NULL THEN 1 ELSE 0 END) AS Missing_CustCity,
    SUM(CASE WHEN CustState IS NULL THEN 1 ELSE 0 END) AS Missing_CustState,
    SUM(CASE WHEN CustZip IS NULL THEN 1 ELSE 0 END) AS Missing_CustZip
```

FROM Customer_Dim;

TABLE_NAME: Customer_Dim	
MISSING_FIRST_NAME	0
MISSING_LAST_NAME	0
MISSING_EMAIL	0
MISSING_PHONE	9
MISSING_CUSTADDRESS	9
MISSING_CUSTCITY	32
MISSING_CUSTSTATE	33
MISSING_CUSTZIP	30

Actions:

- Dropped column Phone & CustAddress, it is not relevant to data insights that we need to analyze.
- Fill Null values with “Unknown” in column City, State and Zip. As these three columns will help us to analyze the Top geographical areas in sales.

Script to handle the missing values.

```
76  --dropping columns phone and address as these are not relevant for our dashboard design.
77  ALTER TABLE CUSTOMER_DIM
78  DROP COLUMN Phone;
79
80  ALTER TABLE CUSTOMER_DIM
81  DROP COLUMN CustAddress;
82
83  -- fill missing values with "Unknown" for CustState, City and Zip.
84  UPDATE Customer_Dim
85  SET CustCity = COALESCE(CustCity, 'Unknown'),
86      CustState = COALESCE(CustState, 'Unknown'),
87      CustZip = COALESCE(CustZip, 'Unknown')
88  WHERE CustCity IS NULL OR CustState IS NULL OR CustZip IS NULL;
89
90  --Rechecking the customer_dim table for missing value.
91  SELECT
92      'Customer_Dim' AS Table_Name,
93      SUM(CASE WHEN First_Name IS NULL THEN 1 ELSE 0 END) AS Missing_First_Name,
94      SUM(CASE WHEN Last_Name IS NULL THEN 1 ELSE 0 END) AS Missing_Last_Name,
95      SUM(CASE WHEN Email IS NULL THEN 1 ELSE 0 END) AS Missing_Email,
96      SUM(CASE WHEN CustCity IS NULL THEN 1 ELSE 0 END) AS Missing_CustCity,
97      SUM(CASE WHEN CustState IS NULL THEN 1 ELSE 0 END) AS Missing_CustState,
98      SUM(CASE WHEN CustZip IS NULL THEN 1 ELSE 0 END) AS Missing_CustZip
99  FROM Customer_Dim;
```

Unique Values in Column CustCity

```
SELECT CustCity,
COUNT(*) AS count
FROM customer_dim
GROUP BY CustCity;
```

Findings – Inconsistency in unique values under the column CustCity.

	CUSTCITY	COUNT
3	Ohio	2
4	TX	22
5	Pennsylvania	7
6	CA	17
7	Colorado	3

Action:

```
106    UPDATE customer_dim
107    SET CustCity = 'Texas'
108    WHERE CustCity IN ('TX');
109
110    UPDATE customer_dim
111    SET CustCity = 'California'
112    WHERE CustCity IN ('CA');
113
114    UPDATE customer_dim
115    SET CustCity = 'Arizona'
116    WHERE CustCity IN ('AZ');
```

Alter Sales/Order Fact table by creating 5 new columns from OrderDt into – Year, Month, Day, Time,AM_PM.

```
-- Creating 3 new columns by splitting ORDERDT into Year, Month, Day, Time and AM_PM.

ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_Year INT;

ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_Month INT;

ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_Day INT;

ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_Time Varchar(8);

ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_AM_PM Varchar(4);
```

```
UPDATE SALES_FACT
SET
    OrderDT_Year = EXTRACT(YEAR FROM OrderDT),
    OrderDT_Month = EXTRACT(MONTH FROM OrderDT),
    OrderDT_Day = EXTRACT(DAY FROM OrderDT),
    OrderDT_Time = TO_CHAR(OrderDT, 'HH24:MI:SS'), -- Format time as HH:MI:SS
    OrderDT_AM_PM = TO_CHAR(OrderDT, 'AM');

ALTER TABLE SALES_FACT
DROP COLUMN ORDERDT;
```

Validating Table after transformation.

	ORIGINAL_ORDERDT	YEAR	MONTH	DAY	TIME	AM_PM
1	2022-07-14 15:43:22.000	2022	7	14	15:43:22	PM
2	2022-09-25 08:25:31.000	2022	9	25	08:25:31	AM
3	2023-07-22 02:54:07.000	2023	7	22	02:54:07	AM
4	2022-03-20 00:00:00.000	2022	3	20	00:00:00	AM
5	2023-10-05 17:16:24.000	2023	10	5	17:16:24	PM
6	2022-03-10 22:21:33.000	2022	3	10	22:21:33	PM
7	2023-04-19 01:44:14.000	2023	4	19	01:44:14	AM
8	2022-04-19 04:37:36.000	2022	4	19	04:37:36	AM
9	2023-02-01 15:28:41.000	2023	2	1	15:28:41	PM
10	2022-12-04 13:41:21.000	2022	12	4	13:41:21	PM
11	2022-02-16 08:14:18.000	2022	2	16	08:14:18	AM
12	2023-02-07 22:56:40.000	2023	2	7	22:56:40	PM
13	2022-07-17 05:40:32.000	2022	7	17	05:40:32	AM
14	2022-04-12 02:57:26.000	2022	4	12	02:57:26	AM
15	2022-01-31 09:47:38.000	2022	1	31	09:47:38	AM
16	2022-05-18 09:06:20.000	2022	5	18	09:06:20	AM
17	2022-03-03 13:11:48.000	2022	3	3	13:11:48	PM
18	2023-04-20 14:56:23.000	2023	4	20	14:56:23	PM
19	2022-04-09 14:36:38.000	2022	4	9	14:36:38	PM
20	2023-09-23 18:11:00.000	2023	9	23	18:11:00	PM

Month Number to Month Names:

```sql
ALTER TABLE SALES_FACT
ADD COLUMN OrderDT_Month_Name VARCHAR(20);

UPDATE SALES_FACT
SET
    OrderDT_Month_Name =
    CASE
        WHEN OrderDT_Month = 1 THEN 'January'
        WHEN OrderDT_Month = 2 THEN 'February'
        WHEN OrderDT_Month = 3 THEN 'March'
        WHEN OrderDT_Month = 4 THEN 'April'
        WHEN OrderDT_Month = 5 THEN 'May'
        WHEN OrderDT_Month = 6 THEN 'June'
        WHEN OrderDT_Month = 7 THEN 'July'
        WHEN OrderDT_Month = 8 THEN 'August'
        WHEN OrderDT_Month = 9 THEN 'September'
        WHEN OrderDT_Month = 10 THEN 'October'
        WHEN OrderDT_Month = 11 THEN 'November'
        WHEN OrderDT_Month = 12 THEN 'December'
        ELSE NULL
    END;
```

Calculated Fields:

```
--Add a DiscountRate column to the Product_dim table
ALTER TABLE PRODUCT_DIM
ADD COLUMN DiscountRate DECIMAL(10, 2);

-- Update the DiscountRate column with the calculated discount rate
UPDATE PRODUCT_DIM
SET DiscountRate = ((MSRP - Price) / MSRP) * 100;

-- Check the updated data
SELECT * FROM PRODUCT_DIM;

ALTER TABLE Product_Dim
ADD COLUMN MarkupPercentage DECIMAL(10, 2);

UPDATE Product_Dim
SET MarkupPercentage = ((Price - Cost) / Cost) * 100;

SELECT * FROM PRODUCT_DIM;
```

Script to copy data from Stage scheme to Prod Schema.

PROJECT_PART2.PROD_DATA Settings

```
1   Create Schema PROJECT_PART2.PROD_DATA;
2
3   create TABLE CUSTOMER_DIM (
4       CUSTOMERID NUMBER(38,0),
5       FIRST_NAME VARCHAR(255),
6       LAST_NAME VARCHAR(255),
7       EMAIL VARCHAR(255),
8       CUSTCITY VARCHAR(100),
9       CUSTSTATE VARCHAR(50),
10      CUSTZIP VARCHAR(50),
11      primary key (CUSTOMERID));
12
13  create TABLE PRODUCT_DIM (
14      PRODUCTID NUMBER(38,0),
15      BRAND VARCHAR(200),
16      NAME VARCHAR(1000),
17      DESCRIPTION VARCHAR(5000),
18      PRICE NUMBER(10,2),
19      MSRP NUMBER(10,2),
20      COST NUMBER(10,2),
21      DISCOUNTRATE NUMBER(10,2),
22      MARKUPPERCENTAGE NUMBER(10,2),
23      primary key (PRODUCTID));
```

```
25  create TABLE SALES_FACT (
26      ORDERID NUMBER(38,0),
27      ORDERAMT NUMBER(12,2),
28      ORDERQTY NUMBER(38,0),
29      CUSTOMERID NUMBER(38,0),
30      ORDERDT_YEAR NUMBER(38,0),
31      ORDERDT_DAY NUMBER(38,0),
32      ORDERDT_TIME TIME(9),
33      ORDERDT_AM_PM VARCHAR(4),
34      TOTAL_SALES DECIMAL(10,2),
35      ORDERDT_MONTH_NAME VARCHAR(20),
36      primary key (ORDERID),
37      foreign key (CUSTOMERID) references CUSTOMER_DIM(CUSTOMERID));
38
39  create TABLE ORDERDETAIL_DIM (
40      ORDERDETAILID NUMBER(38,0),
41      PRODUCTID NUMBER(38,0),
42      ORDERID NUMBER(38,0),
43      primary key (ORDERDETAILID),
44      foreign key (PRODUCTID) references PRODUCT_DIM(PRODUCTID),
45      foreign key (ORDERID) references SALES_FACT(ORDERID));
46
```

```
48  INSERT INTO PROD_DATA.Customer_Dim (CustomerID, First_Name,
49  Last_Name, Email, CustCity, CustState, CustZip)
50  SELECT CustomerID, First_Name, Last_Name,
51  Email, CustCity, CustState, CustZip
52  FROM STAGE_DATA.Customer_Dim;
53
54  INSERT INTO PROD_DATA.PRODUCT_DIM (PRODUCTID, BRAND, NAME,
55  DESCRIPTION, PRICE, MSRP, COST, DISCOUNTRATE, MARKUPPERCENTAGE)
56  SELECT PRODUCTID, BRAND, NAME, DESCRIPTION, PRICE, MSRP, COST,
57  DISCOUNTRATE, MARKUPPERCENTAGE FROM STAGE_DATA.PRODUCT_DIM;
58
59  INSERT INTO PROD_DATA.SALES_FACT (ORDERID, ORDERAMT, ORDERQTY,
60  CUSTOMERID,ORDERDT_YEAR,ORDERDT_DAY, ORDERDT_TIME, ORDERDT_AM_PM,
61  TOTAL_SALES, ORDERDT_MONTH_NAME)
62  SELECT ORDERID, ORDERAMT, ORDERQTY, CUSTOMERID,ORDERDT_YEAR,ORDERDT_DAY,
63  ORDERDT_TIME, ORDERDT_AM_PM, TOTAL_SALES, ORDERDT_MONTH_NAME
64  FROM STAGE_DATA.SALES_FACT;
65
66  INSERT INTO PROD_DATA.ORDERDETAIL_DIM (ORDERDETAILID, PRODUCTID, ORDERID)
67  SELECT ORDERDETAILID, PRODUCTID, ORDERID FROM STAGE_DATA.ORDERDETAIL_DIM;
68
```

Data connected to Tableau Via Snowflake server.