Sql script for Creating schema in staging

```
CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.STAGE_DATA.customer_dim
(
CustomerID number(19,0) not null ,
First_Name varchar(16777216),
Last_Name varchar(16777216),
Email varchar(16777216),
Phone number(10,0),
CustAddress varchar(16777216),
CustState varchar(16777216),
CustCity varchar(16777216),
CustZip varchar(16777216),
primary key (CustomerID)
);

CREATE OR REPLACE TABLE
MSDA3040_TERMPROJECT.STAGE_DATA.orderdetail_dim (
OrderDetailID number(19,0) not null ,
OrderID number(19,0) not null,
ProductID number(19,0) not null,
primary key (OrderDetailID)
);

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.STAGE_DATA.product_dim (
ProductID number(19,0) not null ,
Brand varchar(16777216),
Name varchar(16777216),
description varchar(16777216),
Price float,
MSRP float,
COST float,
primary key (ProductID)
);

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.STAGE_DATA.sales_fact (
OrderID number(19,0),
OrderAmt float,
OrderQty float,
CustomerID number(19,0),
OrderDate DATE,
primary key    (OrderID),
foreign key    (CustomerID)     REFERENCES customer_dim(CustomerID));

--Creating another table for updating city and state from zip code

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.STAGE_DATA.zipcode (
```

```
Customer_ID number(19,0),
State varchar(16777216),
City varchar(16777216),
primary key    (Customer_ID));
```

Sql script for cleaning and transformation of staging data

```
select * from customer_dim ;
select * from sales_fact ;
select * from orderdetail_dim ;
select * from product_dim ;
--based on checking null values only the customer_dim table has null values in 5 columns.
We will be doing more analysis and data cleaning
--on customer_dim

--sales fact tables does not have any null values
--orderdetail table does not have any null values
--productdim table does not have any null values

--cleaning step 1: As phone numbers are important, only 9 rows have null values out of 200
rows so we will keep turn the null values to NA
update  customer_dim
set phone = 'NA'
WHERE phone is null;


select distinct(phone) from customer_dim ;
select * from customer_dim
where CUSTADDRESS is null
and custstate is null
and custcity is null
and custzip is null ;

-- we have onne custom where address, custstate and custzip is null customerid 100 , so we
are going to delete the row.

DELETE FROM customer_dim WHERE customerid =  '100';
-- checking if the customer_id is deleted
select * from customer_dim
where customerid = '100';

select * from customer_dim
where custstate = 'Florida' ;

--we see that certain states have abbreviations in custstate column, this needs to be
changed to full name state
select distinct(custstate) from customer_dim
```

```
where len(custstate)<3;

--updating state abbreviations  to full name, there are 3 states which needs to be updated
update  customer_dim
set custstate = 'Texas'
WHERE custstate = 'TX';

--updating for AZ to Arizona
update  customer_dim
set custstate = 'Arizona'
WHERE custstate = 'AZ';

update  customer_dim
set custstate = 'California'
WHERE custstate = 'CA';

--checking if abbreviations exits after update
select distinct(custstate) from customer_dim
where len(custstate)<3;

--adding zipcodes tables to fill in null values ( this is an experiment to get to fill most of the
null values)
select * from zipcode;

update zipcode
set state = null
where state = ' ';

update zipcode
set city = null
where city = ' ';

select * from customer_dim;


select * from customer_dim
left join zipcode on customer_dim.customerid = zipcode.customer_id;

ALTER TABLE customer_dim
ADD column company_web VARCHAR(1677716);

ALTER TABLE customer_dim
ADD column company VARCHAR(1677716);

select * from customer_dim;

UPDATE customer_dim
SET company_web = split_part(email, '@', -1);
```

```
select * from customer_dim;

UPDATE customer_dim
SET company = split_part(company_web, '.', 1);

ALTER TABLE customer_dim
ADD column State VARCHAR(1677716);

--select * from customer_dim
--left join zipcode on customer_dim.customerid = zipcode.customer_id

ALTER TABLE customer_dim
ADD column city VARCHAR(1677716);

select * from zipcode;


UPDATE customer_dim
SET customer_dim.state = cd.state,
    customer_dim.city = cd.city
FROM zipcode cd
where customer_dim.customerid = cd.customer_id;

select * from customer_dim;

update customer_dim
set state = custstate
where state is null ;

update customer_dim
set city = custcity
where city is null ;

select * from customer_dim;

alter table customer_dim
drop column custstate,custcity;

select * from customer_dim;
```

--we are trying to add data that is required for analysis in the prod schema. This will help us in optimizing for tableau as well
--Assumptions behind this is if a customer has ordered something the address has to be updated, without this the order won't get placed.

```
INSERT INTO MSDA3040_TERMPROJECT.PUBLIC.CUSTOMER_DIM
SELECT *, CURRENT_TIMESTAMP
```

```
FROM STAGE_DATA.CUSTOMER_DIM
where state is not null
and city is not null
and custzip is not null;


INSERT INTO MSDA3040_TERMPROJECT.PUBLIC.sales_fact
SELECT *, CURRENT_TIMESTAMP
FROM STAGE_DATA.sales_fact;

INSERT INTO MSDA3040_TERMPROJECT.PUBLIC.orderdetail_dim
SELECT *, CURRENT_TIMESTAMP
FROM STAGE_DATA.orderdetail_dim;

INSERT INTO MSDA3040_TERMPROJECT.PUBLIC.product_dim
SELECT *, CURRENT_TIMESTAMP
FROM STAGE_DATA.product_dim;
```

SQL script for creating schema and tables in PROD

```
CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.PUBLIC.customer_dim (
CustomerID number(19,0) not null ,
First_Name varchar(16777216),
Last_Name varchar(16777216),
Email varchar(16777216),
Phone varchar(16777216),
CustAddress varchar(16777216),
CustZip varchar(16777216),
Company_web varchar(16777216),
Company varchar(16777216),
State varchar(16777216),
City varchar(16777216),
copy_date TIMESTAMP NOT NULL,
primary key (CustomerID)
);

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.PUBLIC.orderdetail_dim (
OrderDetailID number(19,0) not null ,
OrderID number(19,0) not null,
ProductID number(19,0) not null,
copy_date TIMESTAMP NOT NULL,
primary key (OrderDetailID)
);

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.PUBLIC.product_dim (
ProductID number(19,0) not null ,
Brand varchar(16777216),
```

```
Name varchar(16777216),
description varchar(16777216),
Price float,
MSRP float,
COST float,
copy_date TIMESTAMP NOT NULL,
primary key (ProductID)
);

CREATE OR REPLACE TABLE MSDA3040_TERMPROJECT.PUBLIC.sales_fact (
OrderID number(19,0),
OrderAmt float,
OrderQty float,
CustomerID number(19,0),
OrderDate DATE,
copy_date TIMESTAMP NOT NULL,
primary key    (OrderID),
foreign key    (CustomerID)     REFERENCES customer_dim(CustomerID));
```

During Qc all rows of sales fact,orderdetail,product tables were imported without any transformation. Only the customer dim table was imported truncating it from 200 rows to 161 rows.

```
select * from customer_dim ;
select * from sales_fact ;
select * from orderdetail_dim ;
select * from product_dim ;
```