Olist Ecommerce- Data Management & Analysis

- by Srinivas Abhilash Chintaluru

Table of Contents

Data Source

Dataset: Brazilian e-commerce dataset available in Kaggle

A single dataset is available is normalized into seven datasets/tables as shown below

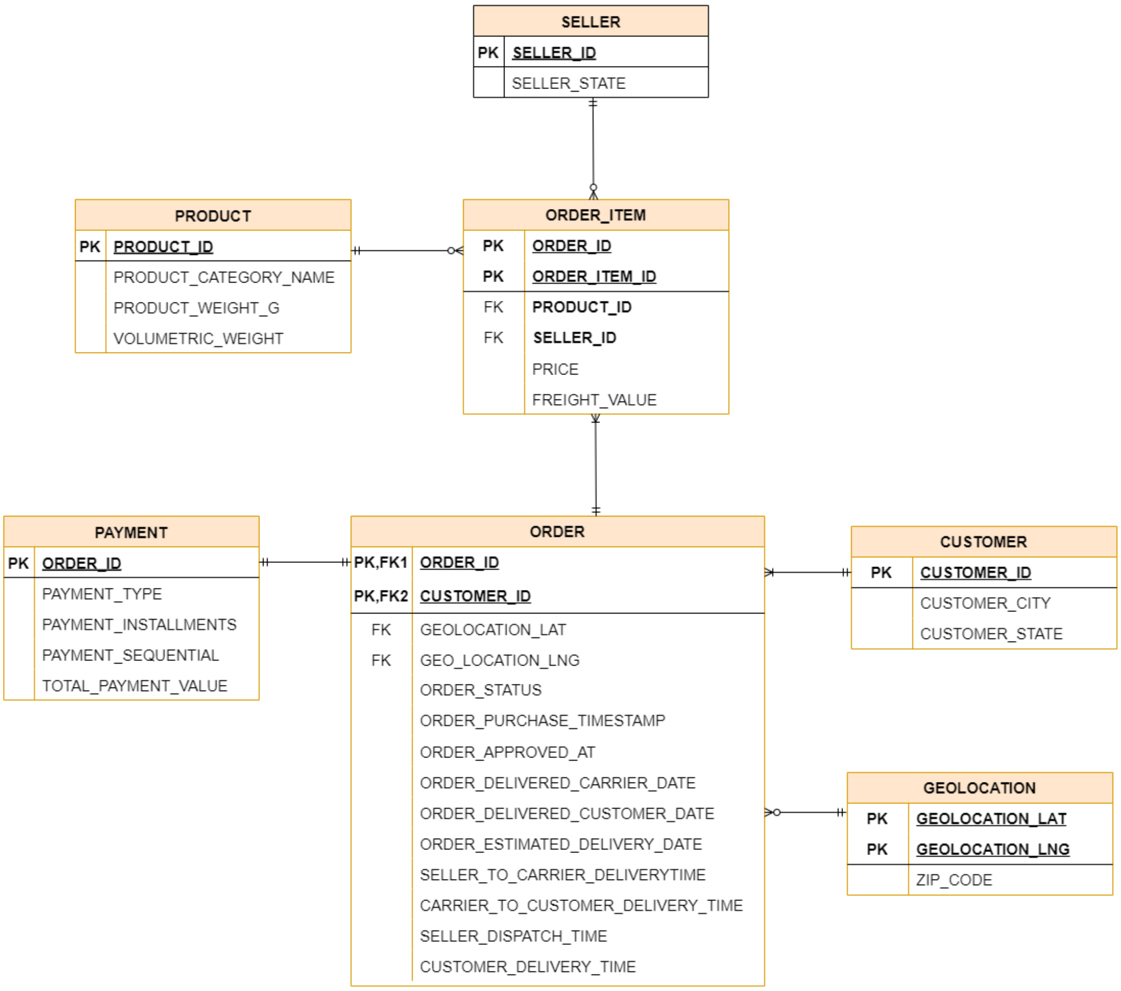
Table name	Table Description
Orders	Information about the orders like order status, order purchase time, order approval time, etc.
Customer	Information about the customers like customer id, city, state, etc..
Product	Information about the product like product id, product category, product weight, etc.
Seller	Information about the seller like seller id, seller state, etc.
Payment	Information about payments like payment type, payment installments, etc.
Delivery	Information related to delivery like the seller to carrier delivery time, carrier to customer delivery time, dispatch time etc.
Geolocation	Information related to latitude and longitude of the orders placed

Entity Relationship Diagram

Relationships used in the ERD:

Definition: Order item – An order can have multiple items in it, each item is assigned one order item id

1) One seller might have sold zero or more order items, and one order item has only one seller
2) One order item has only one product, while one product can be in multiple order items
3) One order can have multiple order items, while one order item belongs to only one order
4) One order has one payment, and one payment has one order
5) One order has only one customer, and one customer can place one more orders
6) One order can be placed from one location, and one location might come up have multiple orders



SELLER	
PK	SELLER_ID
	SELLER_STATE

PRODUCT	
PK	PRODUCT_ID
	PRODUCT_CATEGORY_NAME
	PRODUCT_WEIGHT_G
	VOLUMETRIC_WEIGHT

ORDER_ITEM	
PK	ORDER_ID
PK	ORDER_ITEM_ID
FK	PRODUCT_ID
FK	SELLER_ID
	PRICE
	FREIGHT_VALUE

PAYMENT	
PK	ORDER_ID
	PAYMENT_TYPE
	PAYMENT_INSTALLMENTS
	PAYMENT_SEQUENTIAL
	TOTAL_PAYMENT_VALUE

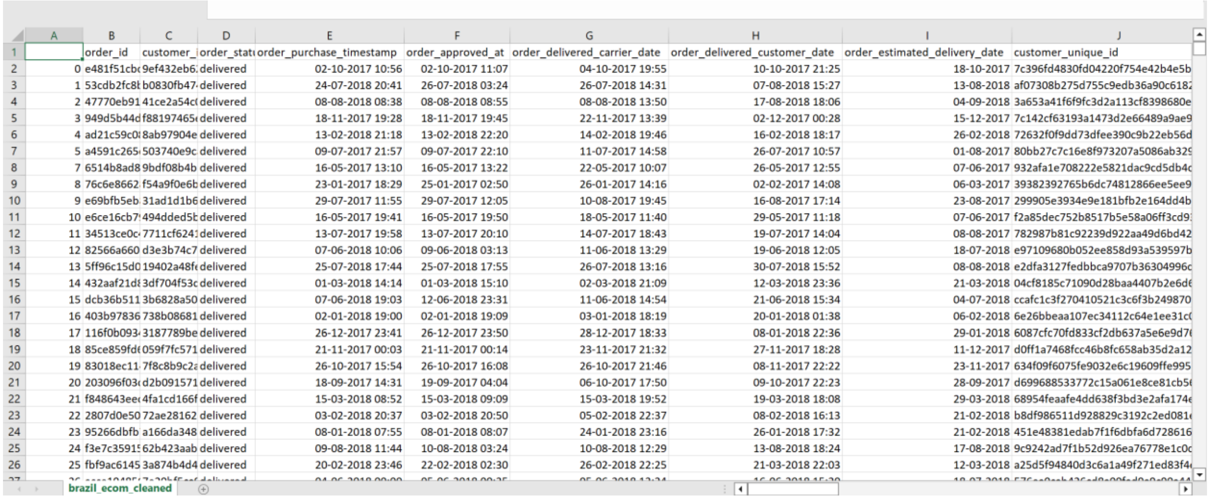
ORDER	
PK,FK1	ORDER_ID
PK,FK2	CUSTOMER_ID
FK	GEOLOCATION_LAT
FK	GEO_LOCATION_LNG
	ORDER_STATUS
	ORDER_PURCHASE_TIMESTAMP
	ORDER_APPROVED_AT
	ORDER_DELIVERED_CARRIER_DATE
	ORDER_DELIVERED_CUSTOMER_DATE
	ORDER_ESTIMATED_DELIVERY_DATE
	SELLER_TO_CARRIER_DELIVERYTIME
	CARRIER_TO_CUSTOMER_DELIVERY_TIME
	SELLER_DISPATCH_TIME
	CUSTOMER_DELIVERY_TIME

CUSTOMER	
PK	CUSTOMER_ID
	CUSTOMER_CITY
	CUSTOMER_STATE

GEOLOCATION	
PK	GEOLOCATION_LAT
PK	GEOLOCATION_LNG
	ZIP_CODE

Database creation, table population, and business questions

My dataset consists of a single CSV file (brazil_ecom_cleaned.csv)



	A	B	C	D	E	F	G	H	I	J
1		order_id	customer_i	order_stat	order_purchase_timestamp	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_estimated_delivery_date	customer_unique_id
2	0	e481f51cbc	9ef432eb6	delivered	02-10-2017 10:56	02-10-2017 11:07	04-10-2017 19:55	10-10-2017 21:25	18-10-2017	7c396fd4830fd04220f754e42b4e5b
3	1	53cdb2fc8b	b0830fb47	delivered	24-07-2018 20:41	26-07-2018 03:24	26-07-2018 14:31	07-08-2018 15:27	13-08-2018	af07308b275d755c9edb36a90c6182
4	2	47770eb91	41ce2a54c	delivered	08-08-2018 08:38	08-08-2018 08:55	08-08-2018 13:50	17-08-2018 18:06	04-09-2018	3a653a41f6f9fc3d2a113cf8398680e
5	3	949d5b44d	f88197465	delivered	18-11-2017 19:28	18-11-2017 19:45	22-11-2017 13:39	02-12-2017 00:28	15-12-2017	7c142cf63193a1473d2e66489a9ae9
6	4	ad21c59c0	8ab97904e	delivered	13-02-2018 21:18	13-02-2018 22:20	14-02-2018 19:46	16-02-2018 18:17	26-02-2018	72632f0f9dd73dfee390c9b22eb56d
7	5	a4591c265	503740e9c	delivered	09-07-2017 21:57	09-07-2017 22:10	11-07-2017 14:58	26-07-2017 10:57	01-08-2017	80bb27c7c16e8f973207a5086ab329
8	7	6514b8ad8	9bdf08b4b	delivered	16-05-2017 13:10	16-05-2017 13:22	22-05-2017 10:07	26-05-2017 12:55	07-06-2017	932afa1e708222e5821dac9cd5db4c
9	8	76c6e8662	f54a9f0e6b	delivered	23-01-2017 18:29	25-01-2017 02:50	26-01-2017 14:16	02-02-2017 14:08	06-03-2017	39382392765b6dc74812866ee5ee9
10	9	e69bfb5eb	31ad1d1b6	delivered	29-07-2017 11:55	29-07-2017 12:05	10-08-2017 19:45	16-08-2017 17:14	23-08-2017	299905e3934e9e181bfb2e164dd4b
11	10	e6ce16cb7	494dded5b	delivered	16-05-2017 19:41	16-05-2017 19:50	18-05-2017 11:40	29-05-2017 11:18	07-06-2017	f2a85dec752b8517b5e58a06ff3cd9
12	11	34513ce0c	7711cf6241	delivered	13-07-2017 19:58	13-07-2017 20:10	14-07-2017 18:43	19-07-2017 14:04	08-08-2017	782987b81c92239d922aa49d6bd42
13	12	82566a660	d3e3b74c7	delivered	07-06-2018 10:06	09-06-2018 03:13	11-06-2018 13:29	19-06-2018 12:05	18-07-2018	e97109680b052ee858d93a539597b
14	13	5ff96c15d0	19402a48f	delivered	25-07-2018 17:44	25-07-2018 17:55	26-07-2018 13:16	30-07-2018 15:52	08-08-2018	e2dfa3127fedbbca9707b36304996c
15	14	432aaf21d	3df704f53c	delivered	01-03-2018 14:14	01-03-2018 15:10	02-03-2018 21:09	12-03-2018 23:36	21-03-2018	04cf8185c71090d28baa4407b2e6d6
16	15	dcb36b511	3b6828a50	delivered	07-06-2018 19:03	12-06-2018 23:31	11-06-2018 14:54	21-06-2018 15:34	04-07-2018	ccafc1c3f270410521c3c6f3b249870
17	16	403b97836	738b08681	delivered	02-01-2018 19:00	02-01-2018 19:09	03-01-2018 18:19	20-01-2018 01:38	06-02-2018	6e26bbeaa107ec34112c64e1ee31c
18	17	116f0b093	3187789be	delivered	26-12-2017 23:41	26-12-2017 23:50	28-12-2017 18:33	08-01-2018 22:36	29-01-2018	6087cfc70fd833cf2db637a5e6e9d7
19	18	85ce859fd	059f7fc571	delivered	21-11-2017 00:03	21-11-2017 00:14	23-11-2017 21:32	27-11-2017 18:28	11-12-2017	d0ff1a7468fcc46b8fc658ab35d2a12
20	19	83018ec11	7f8c8b9c2a	delivered	26-10-2017 15:54	26-10-2017 16:08	26-10-2017 21:46	08-11-2017 22:22	23-11-2017	634f09f6075fe9032e6c19609ffe995
21	20	203096f03	d2b091571	delivered	18-09-2017 14:31	19-09-2017 04:04	06-10-2017 17:50	09-10-2017 22:23	28-09-2017	d699688533772c15a061e8ce81cb5
22	21	f848643ee	4fa1cd166f	delivered	15-03-2018 08:52	15-03-2018 09:09	15-03-2018 19:52	19-03-2018 18:08	29-03-2018	68954feaafe4dd638f3bd3e2afa174
23	22	2807d0e50	72ae28162	delivered	03-02-2018 20:37	03-02-2018 20:50	05-02-2018 22:37	08-02-2018 16:13	21-02-2018	b8df986511d928829c3192c2ed081
24	23	95266dbfb	a166da348	delivered	08-01-2018 07:55	08-01-2018 08:07	24-01-2018 23:16	26-01-2018 17:32	21-02-2018	451e48381edab7f1f6dbfa6d728616
25	24	f3e7c3591	62b423aab	delivered	09-08-2018 11:44	10-08-2018 03:24	10-08-2018 12:29	13-08-2018 18:24	17-08-2018	9c9242ad7f1b52d926ea76778e1c0c
26	25	fbf9ac6145	3a874b4d4	delivered	20-02-2018 23:46	22-02-2018 02:30	26-02-2018 22:25	21-03-2018 22:03	12-03-2018	a25d5f94840d3c6a1a49f271ed83f4

A snippet of the dataset

Approach:

1) Import the dataset into the MySQL database ensuring all the data points have the correct data type.
2) Generate 6 tables using SQL queries
3) Frame business questions, and use SQL queries on the 6 tables to answer the business questions

Database creation and table population

As my dataset is in denormalized form, I created a table for it and loaded the dataset(csv file) into this table.

Step 1: Created a database: ecommerce (Refer to the SQL query in DB_Table_creation.sql)

Step 2: Created a table: ecomm_denom_table in the ecommerce database (Refer to the SQL query in DB_Table_creation.sql)

Step 3: Loading data from brazil_ecom_cleaned.csv file into the ecommerce.ecomm_denom_table, below screenshots illustrate the steps of data loading

Provided the csv file path in the table data import tab

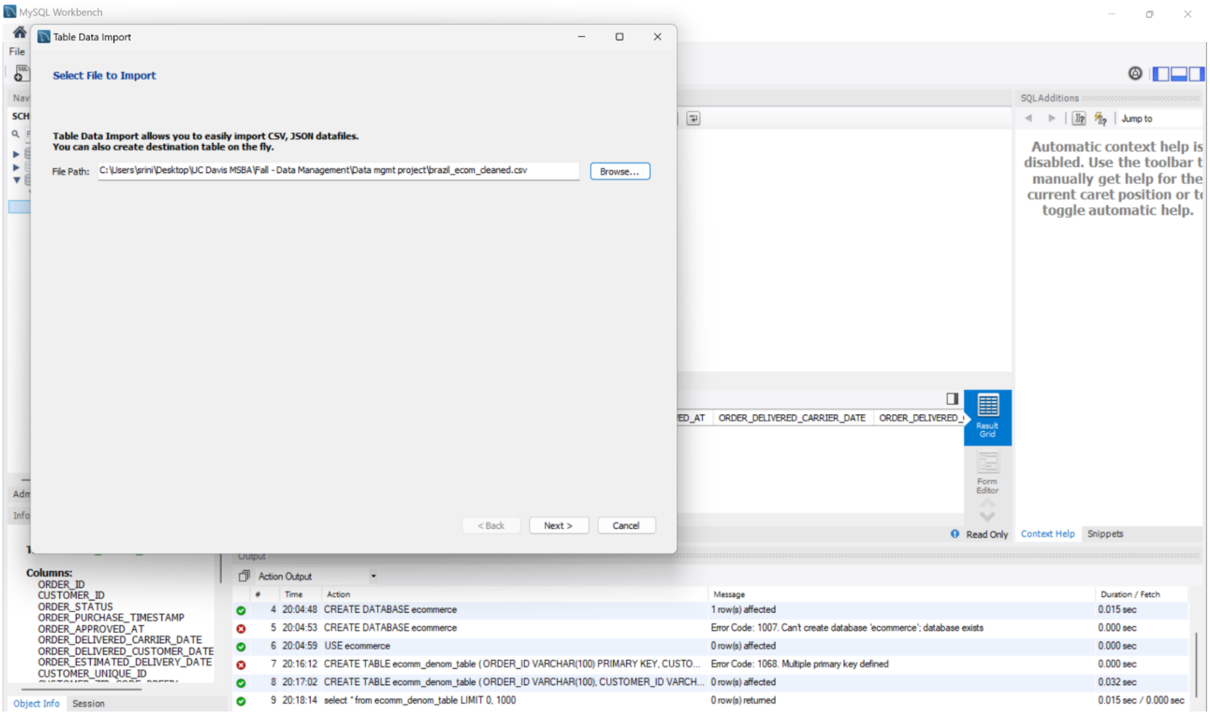


Table Data Import

Select File to Import

Table Data Import allows you to easily import CSV, JSON datafiles.
You can also create destination table on the fly.

File Path: C:\Users\srini\Desktop\UC Davis MSBA\Fall - Data Management\Data mgmt project\brazil_ecom_cleaned.csv Browse...

< Back Next > Cancel

Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

ED_AT | ORDER_DELIVERED_CARRIER_DATE | ORDER_DELIVERED_

Columns:
ORDER_ID
CUSTOMER_ID
ORDER_STATUS
ORDER_PURCHASE_TIMESTAMP
ORDER_APPROVED_AT
ORDER_DELIVERED_CARRIER_DATE
ORDER_DELIVERED_CUSTOMER_DATE
ORDER_ESTIMATED_DELIVERY_DATE
CUSTOMER_UNIQUE_ID

Output

Action Output

	#	Time	Action	Message	Duration / Fetch
✓	4	20:04:48	CREATE DATABASE ecommerce	1 row(s) affected	0.015 sec
✗	5	20:04:53	CREATE DATABASE ecommerce	Error Code: 1007. Can't create database 'ecommerce'; database exists	0.000 sec
✓	6	20:04:59	USE ecommerce	0 row(s) affected	0.000 sec
✗	7	20:16:12	CREATE TABLE ecomm_denom_table (ORDER_ID VARCHAR(100) PRIMARY KEY, CUSTO...	Error Code: 1068. Multiple primary key defined	0.000 sec
✓	8	20:17:02	CREATE TABLE ecomm_denom_table (ORDER_ID VARCHAR(100), CUSTOMER_ID VARCH...	0 row(s) affected	0.032 sec
✓	9	20:18:14	select * from ecomm_denom_table LIMIT 0, 1000	0 row(s) returned	0.015 sec / 0.000 sec

Selected the destination table, in our case it is ecommerce.ecomm_denom_table

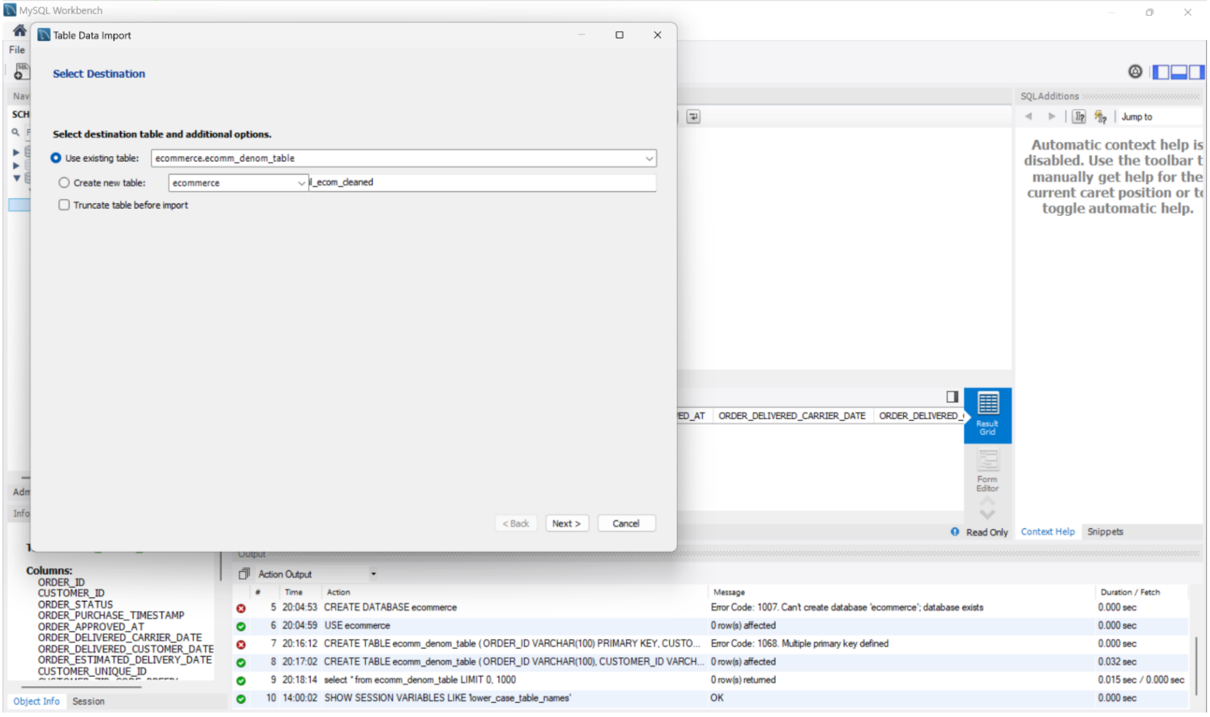


Table Data Import

Select Destination

Select destination table and additional options.

● Use existing table: ecommerce.ecomm_denom_table

○ Create new table: ecommerce l_ecom_cleaned

□ Truncate table before import

< Back Next > Cancel

Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

ED_AT | ORDER_DELIVERED_CARRIER_DATE | ORDER_DELIVERED_

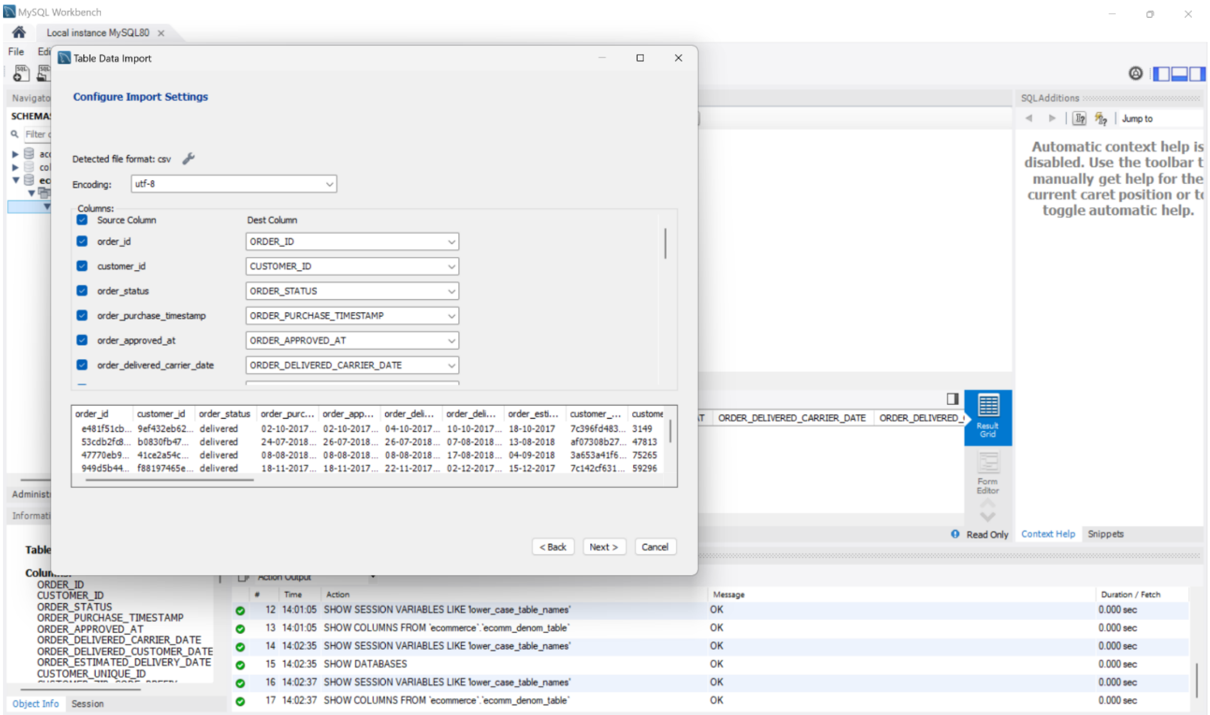
Columns:
ORDER_ID
CUSTOMER_ID
ORDER_STATUS
ORDER_PURCHASE_TIMESTAMP
ORDER_APPROVED_AT
ORDER_DELIVERED_CARRIER_DATE
ORDER_DELIVERED_CUSTOMER_DATE
ORDER_ESTIMATED_DELIVERY_DATE
CUSTOMER_UNIQUE_ID

Output

Action Output

	#	Time	Action	Message	Duration / Fetch
✗	5	20:04:53	CREATE DATABASE ecommerce	Error Code: 1007. Can't create database 'ecommerce'; database exists	0.000 sec
✓	6	20:04:59	USE ecommerce	0 row(s) affected	0.000 sec
✗	7	20:16:12	CREATE TABLE ecomm_denom_table (ORDER_ID VARCHAR(100) PRIMARY KEY, CUSTO...	Error Code: 1068. Multiple primary key defined	0.000 sec
✓	8	20:17:02	CREATE TABLE ecomm_denom_table (ORDER_ID VARCHAR(100), CUSTOMER_ID VARCH...	0 row(s) affected	0.032 sec
✓	9	20:18:14	select * from ecomm_denom_table LIMIT 0, 1000	0 row(s) returned	0.015 sec / 0.000 sec
✓	10	14:00:02	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec

Mapped all the columns in the csv file to the columns created in the table



Detected file format: csv

Encoding: utf-8

Columns:

☑ Source Column	Dest Column
☑ order_id	ORDER_ID
☑ customer_id	CUSTOMER_ID
☑ order_status	ORDER_STATUS
☑ order_purchase_timestamp	ORDER_PURCHASE_TIMESTAMP
☑ order_approved_at	ORDER_APPROVED_AT
☑ order_delivered_carrier_date	ORDER_DELIVERED_CARRIER_DATE

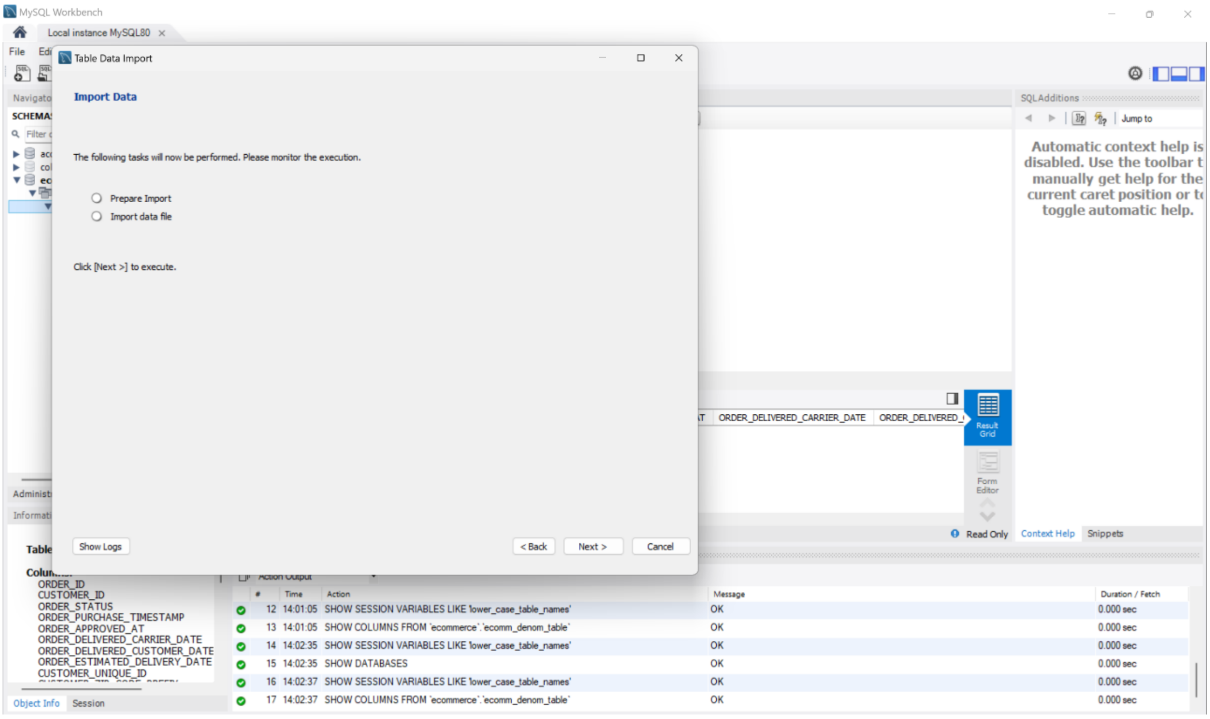
order_id	customer_id	order_status	order_purc...	order_app...	order_deli...	order_deli...	order_esti...	customer_...	custome
e481f51cb...	9ef432eb62...	delivered	02-10-2017...	02-10-2017...	04-10-2017...	10-10-2017...	18-10-2017	7c396fd483...	3149
53cdb2fc8...	b0830fb47...	delivered	24-07-2018...	26-07-2018...	26-07-2018...	07-08-2018...	13-08-2018	af07308b27...	47813
47770eb9...	41ce2a54c...	delivered	08-08-2018...	08-08-2018...	08-08-2018...	17-08-2018...	04-09-2018	3a653a41f6...	75265
949d5b44...	f88197465e...	delivered	18-11-2017...	18-11-2017...	22-11-2017...	02-12-2017...	15-12-2017	7c142cf631...	59296

Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

Columns:
ORDER_ID
CUSTOMER_ID
ORDER_STATUS
ORDER_PURCHASE_TIMESTAMP
ORDER_APPROVED_AT
ORDER_DELIVERED_CARRIER_DATE
ORDER_DELIVERED_CUSTOMER_DATE
ORDER_ESTIMATED_DELIVERY_DATE
CUSTOMER_UNIQUE_ID

#	Time	Action	Message	Duration / Fetch
12	14:01:05	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
13	14:01:05	SHOW COLUMNS FROM `ecommerce`.`ecomm_denom_table`	OK	0.000 sec
14	14:02:35	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
15	14:02:35	SHOW DATABASES	OK	0.000 sec
16	14:02:37	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
17	14:02:37	SHOW COLUMNS FROM `ecommerce`.`ecomm_denom_table`	OK	0.000 sec

Started the import process



The following tasks will now be performed. Please monitor the execution.

○ Prepare Import
○ Import data file

Click [Next >] to execute.

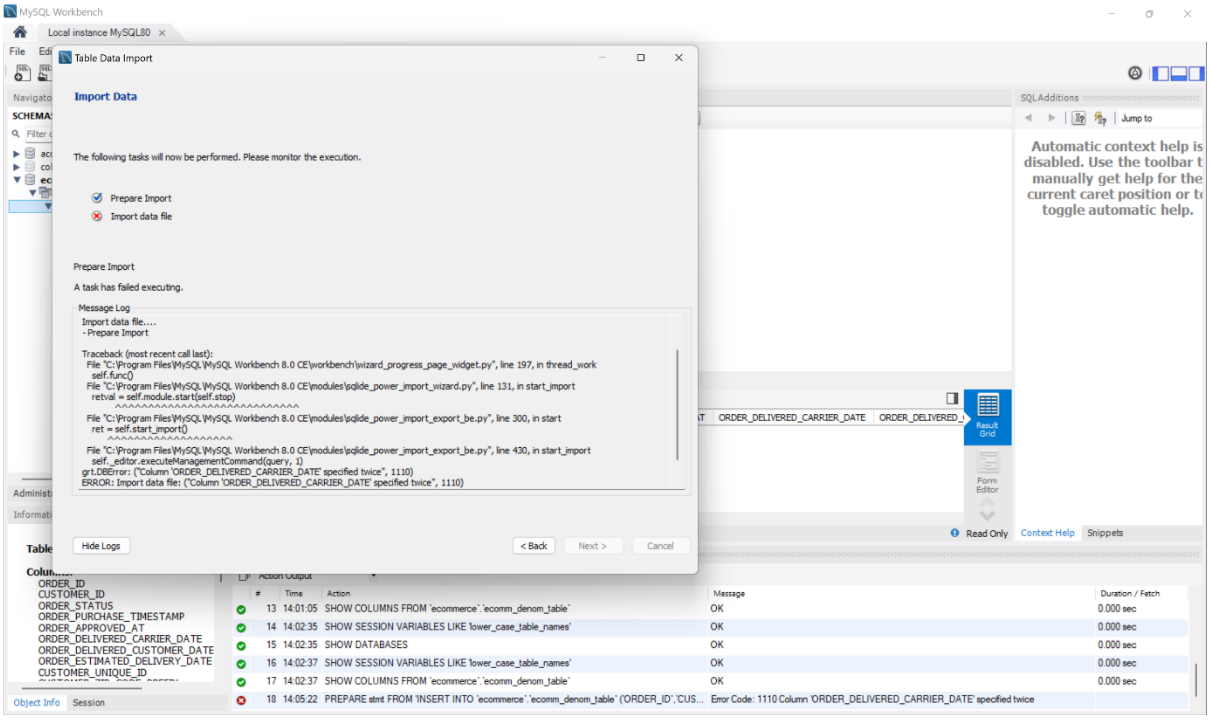
Automatic context help is disabled. Use the toolbar t manually get help for the current caret position or t toggle automatic help.

Columns:
ORDER_ID
CUSTOMER_ID
ORDER_STATUS
ORDER_PURCHASE_TIMESTAMP
ORDER_APPROVED_AT
ORDER_DELIVERED_CARRIER_DATE
ORDER_DELIVERED_CUSTOMER_DATE
ORDER_ESTIMATED_DELIVERY_DATE
CUSTOMER_UNIQUE_ID

#	Time	Action	Message	Duration / Fetch
12	14:01:05	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
13	14:01:05	SHOW COLUMNS FROM `ecommerce`.`ecomm_denom_table`	OK	0.000 sec
14	14:02:35	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
15	14:02:35	SHOW DATABASES	OK	0.000 sec
16	14:02:37	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
17	14:02:37	SHOW COLUMNS FROM `ecommerce`.`ecomm_denom_table`	OK	0.000 sec

Data import Error 1:

Received an error as the column appeared twice



The following tasks will now be performed. Please monitor the execution.

☑ Prepare Import
☒ Import data file

Prepare Import

A task has failed executing.

Message Log

```
Import data file....
- Prepare Import

Traceback (most recent call last):
  File "C:\Program Files\MySQL\MySQL Workbench 8.0 CE\workbench\wizard_progress_page_widget.py", line 197, in thread_work
    self.func()
  File "C:\Program Files\MySQL\MySQL Workbench 8.0 CE\modules\sqlide_power_import_wizard.py", line 131, in start_import
    retval = self.module.start(self.stop)
             ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
  File "C:\Program Files\MySQL\MySQL Workbench 8.0 CE\modules\sqlide_power_import_export_be.py", line 300, in start
    ret = self.start_import()
          ^^^^^^^^^^^^^^^^^^^
  File "C:\Program Files\MySQL\MySQL Workbench 8.0 CE\modules\sqlide_power_import_export_be.py", line 430, in start_import
    self._editor.executeManagementCommand(query, 1)
grt.DBError: ("Column 'ORDER_DELIVERED_CARRIER_DATE' specified twice", 1110)
ERROR: Import data file: ("Column 'ORDER_DELIVERED_CARRIER_DATE' specified twice", 1110)
```

Columns:
ORDER_ID
CUSTOMER_ID
ORDER_STATUS
ORDER_PURCHASE_TIMESTAMP
ORDER_APPROVED_AT
ORDER_DELIVERED_CARRIER_DATE
ORDER_DELIVERED_CUSTOMER_DATE
ORDER_ESTIMATED_DELIVERY_DATE
CUSTOMER_UNIQUE_ID

#	Time	Action	Message	Duration / Fetch
13	14:01:05	SHOW COLUMNS FROM \`ecommerce\`.\`ecomm_denom_table\`	OK	0.000 sec
14	14:02:35	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
15	14:02:35	SHOW DATABASES	OK	0.000 sec
16	14:02:37	SHOW SESSION VARIABLES LIKE 'lower_case_table_names'	OK	0.000 sec
17	14:02:37	SHOW COLUMNS FROM \`ecommerce\`.\`ecomm_denom_table\`	OK	0.000 sec
18	14:05:22	PREPARE stmt FROM 'INSERT INTO \`ecommerce\`.\`ecomm_denom_table\` (\`ORDER_ID\`,\`CUS...	Error Code: 1110 Column 'ORDER_DELIVERED_CARRIER_DATE' specified twice	

As the column 'ORDER_DELIVERED_CARRIER_DATE' was repeating twice, I went to the previous step of mapping the columns and fixed it.

Data import Error 2:

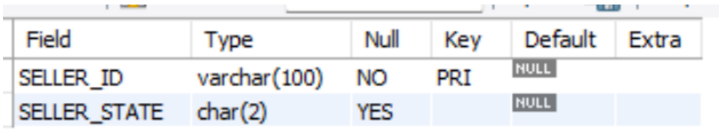
After fixing this, I received another error due to the issue in the date format of date-based columns in the CSV file. Changed the date format from 'dd-mm-yyyy HH:MM:SS' to 'yyyy-mm-dd HH:MM:SS' and fixed the issue.

Step 4: Created 7 tables: seller, product, customer, payment, geolocation, orders, order_item (Refer to the SQL query in DB_Table_creation.sql)

Step 5: Load data from the denormalized table into each of the 7 tables created

Data dictionary

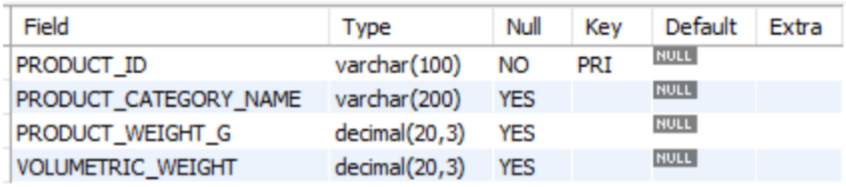
SELLER table



Field	Type	Null	Key	Default	Extra
SELLER_ID	varchar(100)	NO	PRI	NULL	
SELLER_STATE	char(2)	YES		NULL	

Field	Data Description
SELLER_ID	The unique ID for each seller
SELLER_STATE	The state that the seller belongs to

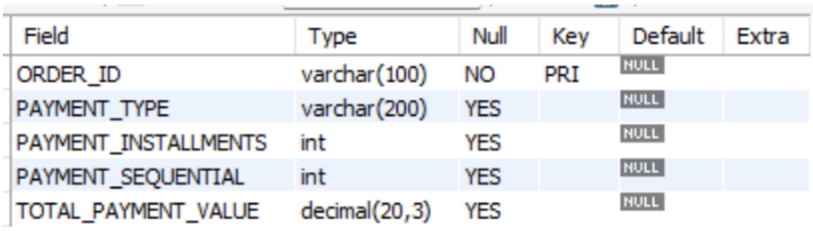
PRODUCT table



Field	Type	Null	Key	Default	Extra
PRODUCT_ID	varchar(100)	NO	PRI	NULL	
PRODUCT_CATEGORY_NAME	varchar(200)	YES		NULL	
PRODUCT_WEIGHT_G	decimal(20,3)	YES		NULL	
VOLUMETRIC_WEIGHT	decimal(20,3)	YES		NULL	

Field	Data Description
PRODUCT_ID	The unique ID for each product
PRODUCT_CATEGORY_NAME	Category of the product
PRODUCT_WEIGHT_G	Weight of the product
VOLUMETRIC_WEIGHT	Volume of the product ordered

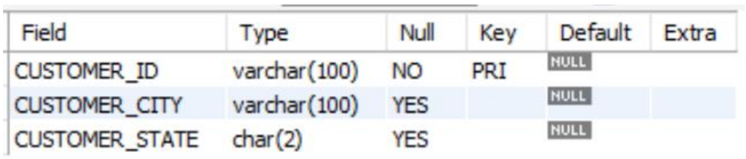
PAYMENT table



Field	Type	Null	Key	Default	Extra
ORDER_ID	varchar(100)	NO	PRI	NULL	
PAYMENT_TYPE	varchar(200)	YES		NULL	
PAYMENT_INSTALLMENTS	int	YES		NULL	
PAYMENT_SEQUENTIAL	int	YES		NULL	
TOTAL_PAYMENT_VALUE	decimal(20,3)	YES		NULL	

Field	Data Description
PRODUCT_ID	The unique ID for each product
PRODUCT_CATEGORY_NAME	Category of the product
PRODUCT_WEIGHT_G	Weight of the product
VOLUMETRIC_WEIGHT	Volume of the product ordered

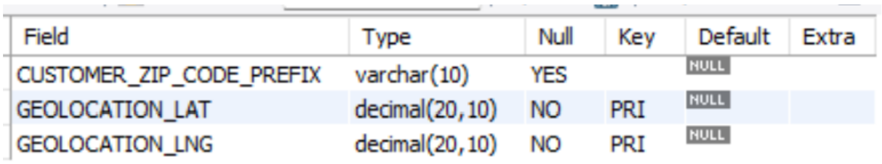
CUSTOMER table



Field	Type	Null	Key	Default	Extra
CUSTOMER_ID	varchar(100)	NO	PRI	NULL	
CUSTOMER_CITY	varchar(100)	YES		NULL	
CUSTOMER_STATE	char(2)	YES		NULL	

Field	Data Description
CUSTOMER_ID	The unique ID for each customer
CUSTOMER_CITY	City of the customer
CUSTOMER_STATE	State of the customer

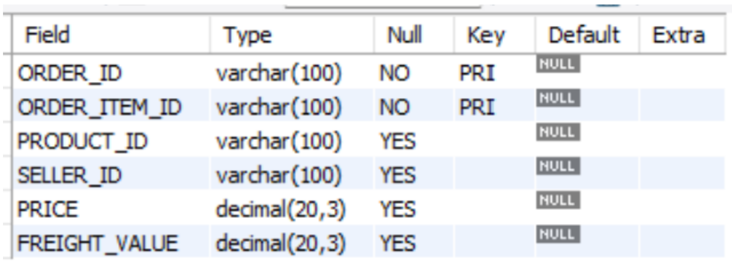
GEOLOCATION table



Field	Type	Null	Key	Default	Extra
CUSTOMER_ZIP_CODE_PREFIX	varchar(10)	YES		NULL	
GEOLOCATION_LAT	decimal(20,10)	NO	PRI	NULL	
GEOLOCATION_LNG	decimal(20,10)	NO	PRI	NULL	

Field	Data Description
CUSTOMER_ZIP_CODE_PREFIX	The unique ID for each customer
GEOLOCATION_LAT	City of the customer
CUSTOMER_STATE	State of the customer

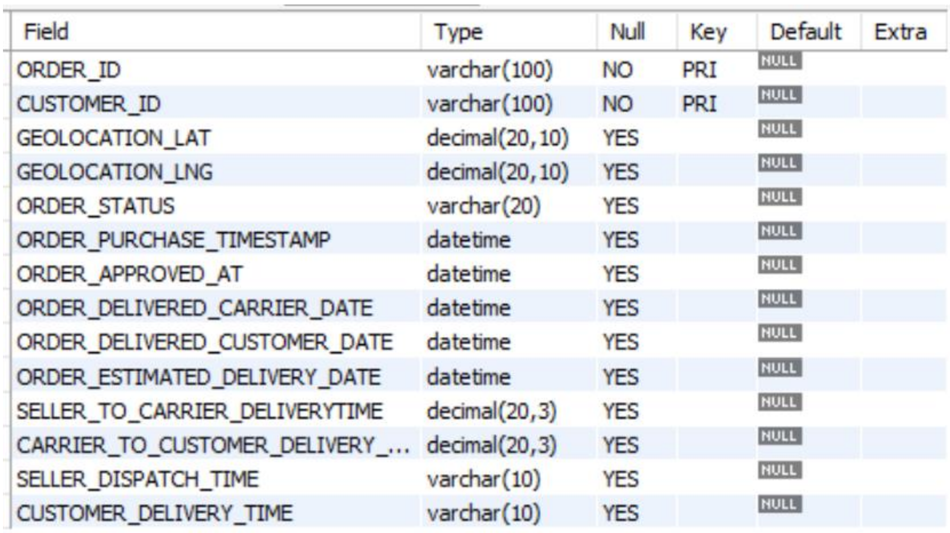
ORDER_ITEM table



Field	Type	Null	Key	Default	Extra
ORDER_ID	varchar(100)	NO	PRI	NULL	
ORDER_ITEM_ID	varchar(100)	NO	PRI	NULL	
PRODUCT_ID	varchar(100)	YES		NULL	
SELLER_ID	varchar(100)	YES		NULL	
PRICE	decimal(20,3)	YES		NULL	
FREIGHT_VALUE	decimal(20,3)	YES		NULL	

Field	Data Description
ORDER_ID	The unique ID for each order
ORDER_ITEM_ID	The unique ID for each item in the order
PRODUCT_ID	The unique ID for the product
SELLER_ID	The unique ID of the seller
PRICE	Price of the product
FREIGHT_VALUE	Delivery value of the order

ORDERS table

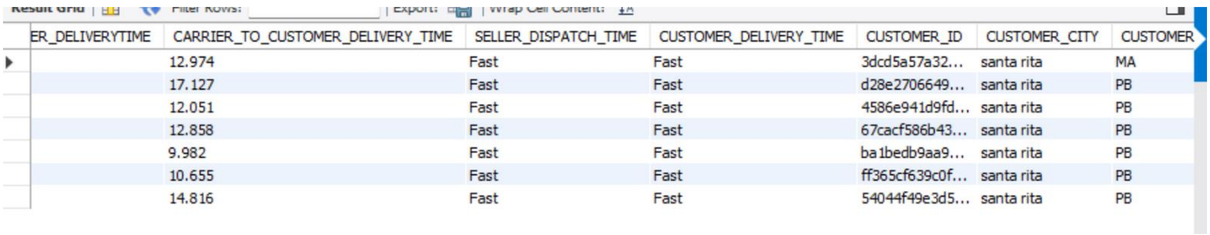


Field	Type	Null	Key	Default	Extra
ORDER_ID	varchar(100)	NO	PRI	NULL	
CUSTOMER_ID	varchar(100)	NO	PRI	NULL	
GEOLOCATION_LAT	decimal(20,10)	YES		NULL	
GEOLOCATION_LNG	decimal(20,10)	YES		NULL	
ORDER_STATUS	varchar(20)	YES		NULL	
ORDER_PURCHASE_TIMESTAMP	datetime	YES		NULL	
ORDER_APPROVED_AT	datetime	YES		NULL	
ORDER_DELIVERED_CARRIER_DATE	datetime	YES		NULL	
ORDER_DELIVERED_CUSTOMER_DATE	datetime	YES		NULL	
ORDER_ESTIMATED_DELIVERY_DATE	datetime	YES		NULL	
SELLER_TO_CARRIER_DELIVERYTIME	decimal(20,3)	YES		NULL	
CARRIER_TO_CUSTOMER_DELIVERY_...	decimal(20,3)	YES		NULL	
SELLER_DISPATCH_TIME	varchar(10)	YES		NULL	
CUSTOMER_DELIVERY_TIME	varchar(10)	YES		NULL	

Field	Data Description
ORDER_ID	The unique ID for each order
CUSTOMER_ID	The unique ID for each customer
GEOLOCATION_LAT	Latitude of the order location
GEOLOCATION_LNG	Longitude of the order location
ORDER_STATUS	Status of the order (delivered or cancelled)
ORDER_PURCHASE_TIMESTAMP	The timestamp of the order purchase
ORDER_APPROVED_AT	The timestamp of the order approval
ORDER_DELIVERED_CARRIER_DATE	The timestamp of order delivery to the carrier
ORDER_DELIVERED_CUSTOMER_DATE	The timestamp of order delivery to the customer
ORDER_ESTIMATED_DELIVERY_DATE	The estimated delivery date to the customer
SELLER_TO_CARRIER_DELIVERYTIME	Time taken to reach from seller to carrier
CARRIER_TO_CUSTOMER_DELIVERYTIME	Time taken to reach from carrier to customer
SELLER_DISPATCH_TIME	Time taken by the seller to dispatch (binary: delay or fast)
CUSTOMER_DELIVERY_TIME	Time taken for the delivery (binary: delay or fast)

Business Hypothesis & Analysis

1. The e-commerce company started receiving complaints from the city ‘santa rita’. The management wanted to have a look at the details of the orders from ‘santa rita’ and understand if there are any issues after looking at the data.



ER_DELIVERYTIME	CARRIER_TO_CUSTOMER_DELIVERY_TIME	SELLER_DISPATCH_TIME	CUSTOMER_DELIVERY_TIME	CUSTOMER_ID	CUSTOMER_CITY	CUSTOMER
	12.974	Fast	Fast	3dcd5a57a32...	santa rita	MA
	17.127	Fast	Fast	d28e2706649...	santa rita	PB
	12.051	Fast	Fast	4586e941d9fd...	santa rita	PB
	12.858	Fast	Fast	67cacf586b43...	santa rita	PB
	9.982	Fast	Fast	ba1bedb9aa9...	santa rita	PB
	10.655	Fast	Fast	ff365cf639c0f...	santa rita	PB
	14.816	Fast	Fast	54044f49e3d5...	santa rita	PB

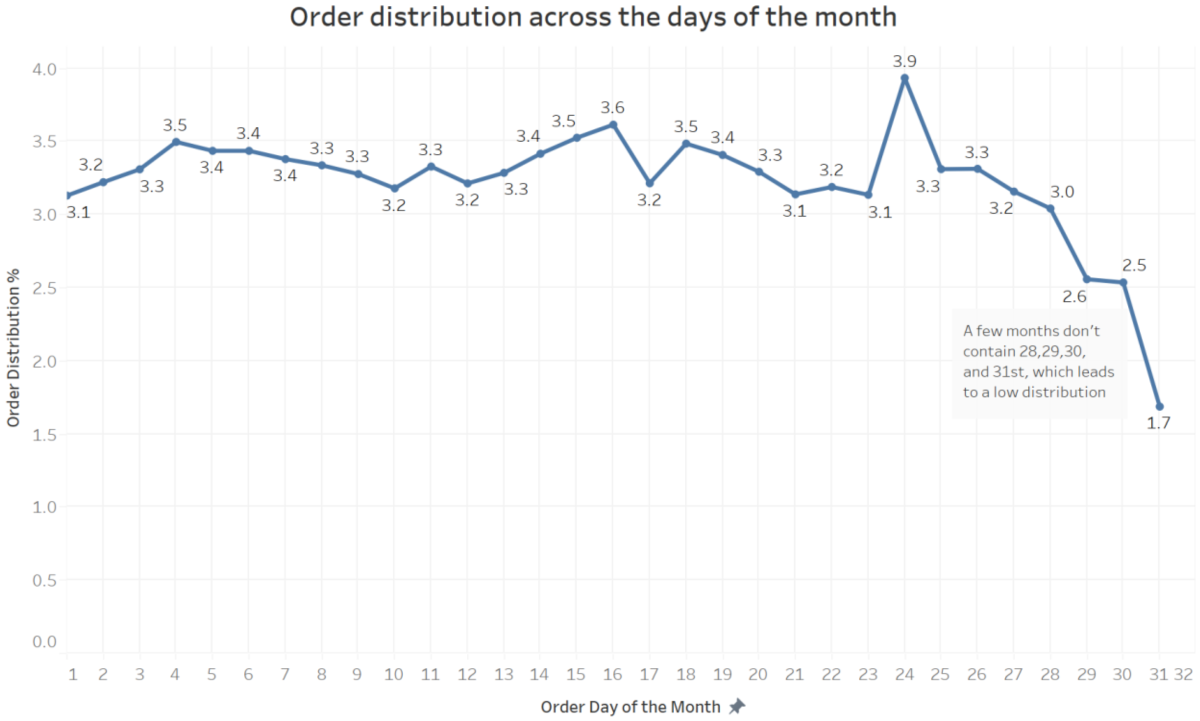
There were a 7 orders from santa rita, the details of which were presented to the management.

2. What is the average order value and average delivery cost?

Avg. order value	Avg. delivery price
R$ 125.3	R$ 20.6

Avg. delivery price is approximately one-sixth of the avg. order value

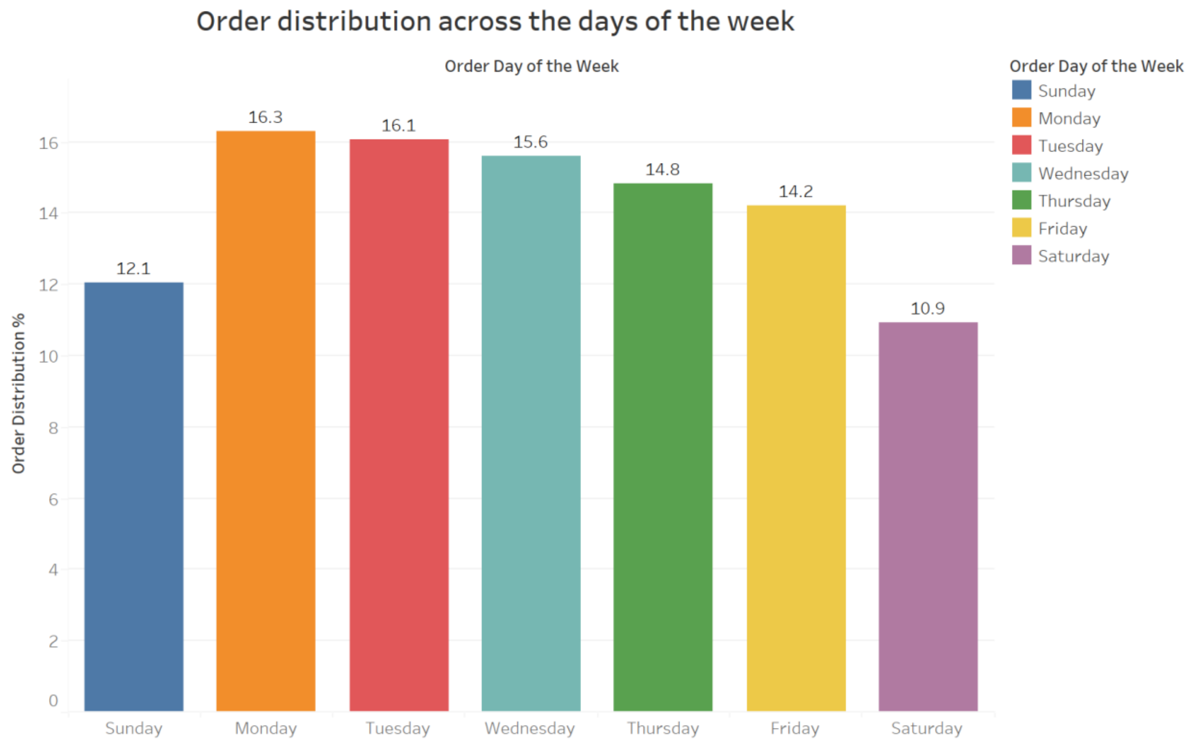
3. How are the orders distributed over the days of the month? and days of the week?



A few months don’t contain 28,29,30, and 31st, which leads to a low distribution

The trend of sum of Order Distribution for Order Day Month. The marks are labelled by sum of Order Distribution.

The number of orders peak on 16th and 24th dates of the month, while 1st, 21st, 23rd and 28th record low orders.



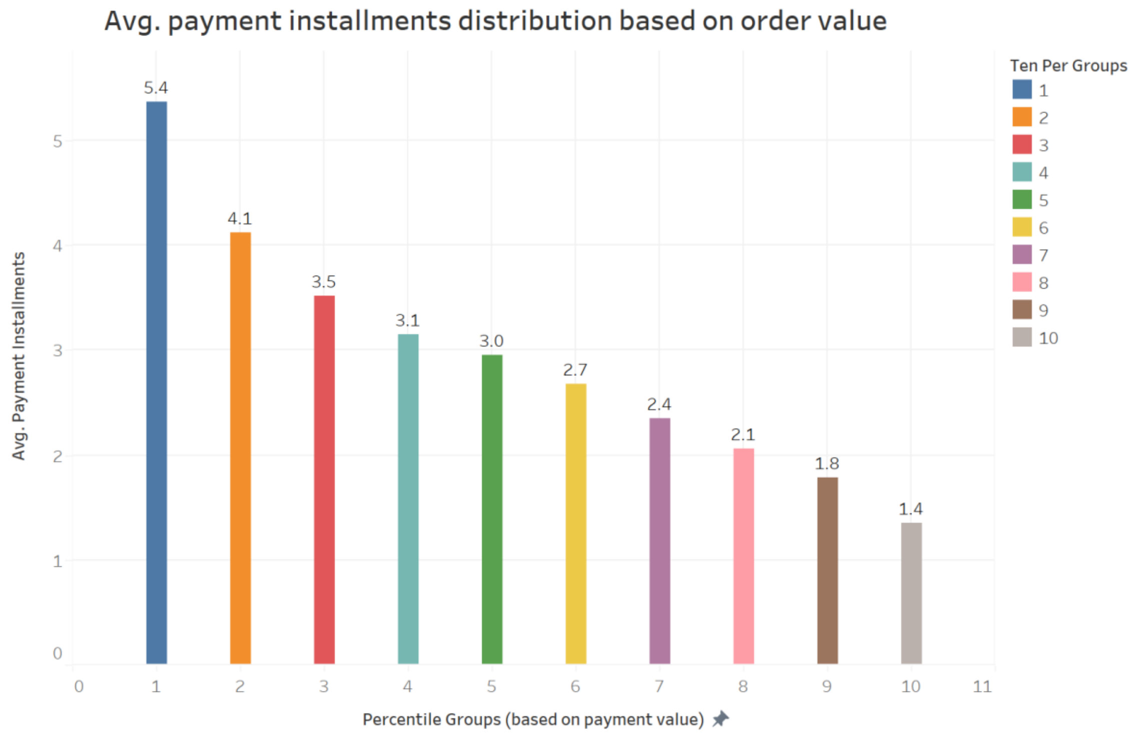
Sum of Order Distribution for each Order Day of the Week. Colour shows details about Order Day of the Week. The marks are labelled by sum of Order Distribution.

Mondays and Tuesdays have higher % of orders, while Saturdays and Sundays have a significantly lower % of orders

4. What are the highest and lowest ordered product categories?

Product Categories	
Highest order	**Lowest ordered**
Bed, Table and Bath products	Kitchen
Beauty and Health care	Gaming PCs
Sport	Children clothing
Computer accessories	Insurance and Services

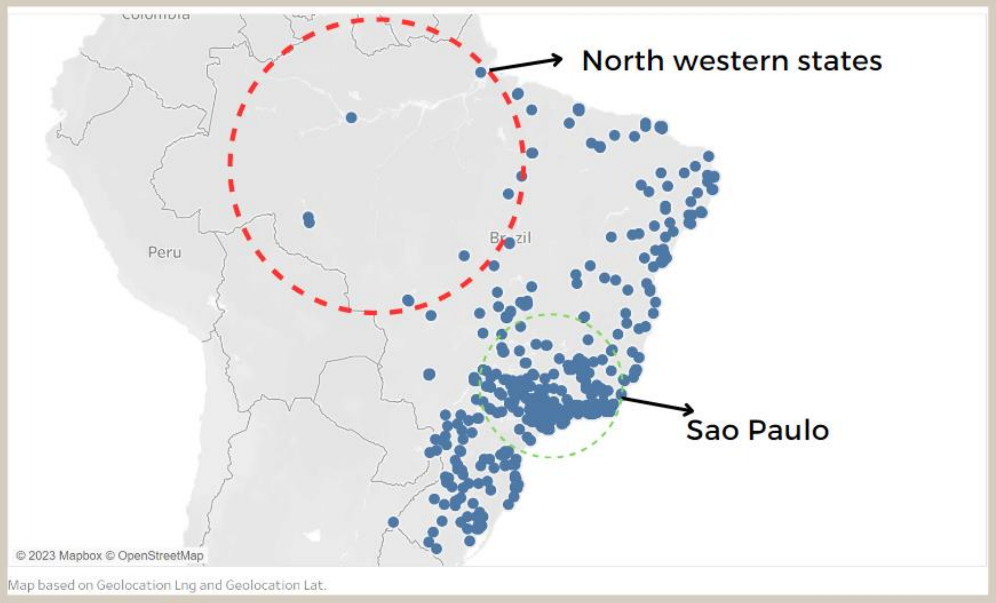
5. Do customers prefer more installments for higher-value products?



The plot of average of Avg Payment Installments for Ten Per Groups. Colour shows details about Ten Per Groups. The marks are labelled by sum of Avg Payment Installments.

The average payment installments decreases as the order value increases

6. The management is interested in knowing the states with the highest and the lowest orders as it would help allocate the workforce accordingly. The distribution of orders across different states in the country. A detailed heatmap would help the organization allocate the workforce more efficiently.



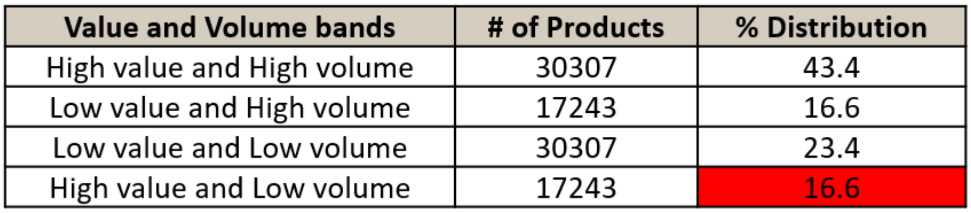
Map based on Geolocation Lng and Geolocation Lat.

Sao Paulo state has the highest frequency of orders, while most of the north western regions have no orders

7. What is the percentage of orders that are dispatched late by the seller (seller_dispatch_time = Delay) but delivered before the promised time to the customer (customer_delivery_time = Fast)?

6.8% of the orders are dispatched late by the sellers but received on time by the customers

8. As high-value and low-volume products generate a lot of revenue for an e-commerce company, What is the distribution of (high-value, low-volume), (low-value, high-volume), (low-value, low-volume), (high-value, high-volume) products are delivered before the promised time to the customer?



Value and Volume bands	# of Products	% Distribution
High value and High volume	30307	43.4
Low value and High volume	17243	16.6
Low value and Low volume	30307	23.4
High value and Low volume	17243	16.6

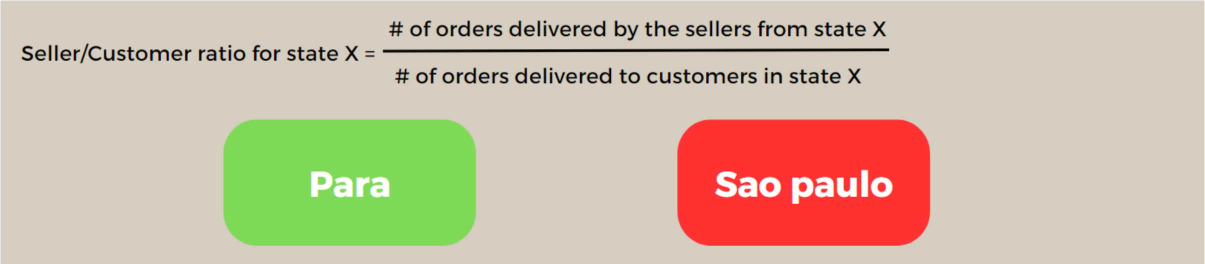
The distribution of High value and Low volume is quite less as compared to the counterparts

9. How many orders in total are using more than 5 vouchers to pay the order amount?

124

The number of orders with more than 5 vouchers are quite less

10. As it would decrease the delivery costs, the management wants to run marketing campaigns and generate more orders from the states where we have sellers but very few customers.



$$\text{Seller/Customer ratio for state X} = \frac{\text{\# of orders delivered by the sellers from state X}}{\text{\# of orders delivered to customers in state X}}$$

Para

Sao paulo

11. The management would like to run a credit card campaign if there are a significant number of customers who are not using credit cards to pay the orders, what is the percentage of customers who are using a credit card for a full payment/ partial payment?

77%

The % of orders using credit cards for partial/complete payment are quite high