

```
In [2]: #Imported the Required Libraries
import pandas as pd
import numpy as np
```

```
In [3]: # reading the raw csv files
data = pd.read_csv("D:\Python_sql project\orders.csv")
```

```
In [5]: # Printing the first 10 rows from the data
data.head(10)
```

```
Out[5]:
```

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	Discount Percent
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240	260	2	2
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600	730	3	3
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10	10	2	5
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780	960	5	2
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20	20	2	5
5	6	2022-03-13	Not Available	Consumer	United States	Los Angeles	California	90032	West	Furniture	Furnishings	FUR-FU-10001487	50	50	7	3
6	7	2022-12-28	Standard Class	Consumer	United States	Los Angeles	California	90032	West	Office Supplies	Art	OFF-AR-10002833	10	10	4	3
7	8	2022-01-25	Standard Class	Consumer	United States	Los Angeles	California	90032	West	Technology	Phones	TEC-PH-10002275	860	910	6	5
8	9	2023-03-23	Not Available	Consumer	United States	Los Angeles	California	90032	West	Office Supplies	Binders	OFF-BI-10003910	20	20	3	2
9	10	2023-05-16	Standard Class	Consumer	United States	Los Angeles	California	90032	West	Office Supplies	Appliances	OFF-AP-10002892	90	110	5	3

```
In [6]: # Printing the last 10 rows from the data
data.tail(10)
```

Out[6]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	Disc Pe
9984	9985	2023-08-04	Standard Class	Consumer	United States	Long Beach	New York	11561	East	Office Supplies	Labels	OFF-LA-10003766	30	30	10	
9985	9986	2023-09-30	Standard Class	Consumer	United States	Long Beach	New York	11561	East	Office Supplies	Supplies	OFF-SU-10000898	50	60	4	
9986	9987	2022-11-30	Standard Class	Consumer	United States	Los Angeles	California	90008	West	Technology	Accessories	TEC-AC-10003399	40	40	1	
9987	9988	2023-04-01	Standard Class	Corporate	United States	Athens	Georgia	30605	South	Technology	Accessories	TEC-AC-10001539	80	80	1	
9988	9989	2022-11-17	Standard Class	Corporate	United States	Athens	Georgia	30605	South	Technology	Phones	TEC-PH-10004006	180	210	5	
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	FUR-FU-10001889	30	30	3	
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	FUR-FU-10000747	70	90	2	
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	TEC-PH-10003645	220	260	2	
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	OFF-PA-10004041	30	30	4	
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	OFF-AP-10002684	210	240	2	

```
In [7]: # read data from the file and handle null values by using na_values
data = pd.read_csv("D:\Python_sql project\orders.csv", na_values=['Not Available','unknown'])
data.tail(50)
```

Out[7]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	I
9944	9945	2022-09-25	Second Class	Home Office	United States	Seattle	Washington	98103	West	Office Supplies	Storage	OFF-ST-10000419	30	40	3	
9945	9946	2023-06-26	Standard Class	Corporate	United States	Philadelphia	Pennsylvania	19120	East	Office Supplies	Paper	OFF-PA-10000327	0	0	1	
9946	9947	2022-04-16	Standard Class	Corporate	United States	Philadelphia	Pennsylvania	19120	East	Technology	Accessories	TEC-AC-10004353	120	150	3	
9947	9948	2022-10-31	Second Class	Corporate	United States	Indianapolis	Indiana	46203	Central	Furniture	Chairs	FUR-CH-10003746	1600	1930	6	
9948	9949	2022-09-13	Second Class	Corporate	United States	Indianapolis	Indiana	46203	Central	Office Supplies	Appliances	OFF-AP-10002945	1950	2410	8	
9949	9950	2022-01-07	Second Class	Corporate	United States	Indianapolis	Indiana	46203	Central	Technology	Accessories	TEC-AC-10004568	70	80	3	
9950	9951	2022-03-19	Second Class	Corporate	United States	Indianapolis	Indiana	46203	Central	Technology	Accessories	TEC-AC-10001714	40	40	1	
9951	9952	2022-04-07	Second Class	Corporate	United States	Indianapolis	Indiana	46203	Central	Office Supplies	Binders	OFF-BI-10002072	20	20	2	
9952	9953	2022-09-28	Second Class	Corporate	United States	Los Angeles	California	90045	West	Office Supplies	Binders	OFF-BI-10001153	50	60	2	
9953	9954	2022-10-27	Second Class	Corporate	United States	Los Angeles	California	90045	West	Office Supplies	Paper	OFF-PA-10004983	10	10	1	
9954	9955	2023-08-02	Second Class	Corporate	United States	Los Angeles	California	90045	West	Office Supplies	Binders	OFF-BI-10000948	20	30	3	
9955	9956	2022-03-18	Second Class	Corporate	United States	Los Angeles	California	90045	West	Furniture	Tables	FUR-TA-10002622	220	270	2	
9956	9957	2022-10-10	Standard Class	Home Office	United States	New Rochelle	New York	10801	East	Office Supplies	Paper	OFF-PA-10001776	50	50	5	
9957	9958	2023-01-03	Standard Class	Home Office	United States	New Rochelle	New York	10801	East	Office Supplies	Paper	OFF-PA-10004039	210	220	4	
9958	9959	2023-12-21	Standard Class	Home Office	United States	New Rochelle	New York	10801	East	Office Supplies	Supplies	OFF-SU-10001574	10	10	2	

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	I
9959	9960	2023-02-09	Standard Class	Consumer	United States	Chandler	Arizona	85224	West	Office Supplies	Art	OFF-AR-10001761	10	10	2	
9960	9961	2023-02-26	Second Class	Home Office	United States	Florence	Kentucky	41042	South	Technology	Accessories	TEC-AC-10002305	20	20	1	
9961	9962	2022-04-18	First Class	Home Office	United States	Houston	Texas	77041	Central	Office Supplies	Paper	OFF-PA-10000675	60	70	2	
9962	9963	2022-09-02	First Class	Home Office	United States	Houston	Texas	77041	Central	Furniture	Bookcases	FUR-BO-10004218	320	380	4	
9963	9964	2022-01-17	Same Day	Consumer	United States	Philadelphia	Pennsylvania	19140	East	Office Supplies	Paper	OFF-PA-10003072	10	10	2	
9964	9965	2023-07-08	Second Class	Corporate	United States	Newark	Delaware	19711	East	Furniture	Furnishings	FUR-FU-10002671	10	10	1	
9965	9966	2022-12-31	Second Class	Corporate	United States	Newark	Delaware	19711	East	Office Supplies	Paper	OFF-PA-10000349	0	0	1	
9966	9967	2022-02-21	Second Class	Corporate	United States	Newark	Delaware	19711	East	Office Supplies	Envelopes	OFF-EN-10004483	100	110	7	
9967	9968	2022-10-02	Standard Class	Consumer	United States	Plainfield	New Jersey	7060	East	Office Supplies	Binders	OFF-BI-10004209	30	40	5	
9968	9969	2023-04-15	Standard Class	Consumer	United States	Plainfield	New Jersey	7060	East	Office Supplies	Binders	OFF-BI-10004600	700	740	2	
9969	9970	2023-11-18	Standard Class	Consumer	United States	Plainfield	New Jersey	7060	East	Office Supplies	Appliances	OFF-AP-10003622	20	20	7	
9970	9971	2023-05-04	Standard Class	Home Office	United States	Smyrna	Georgia	30080	South	Office Supplies	Binders	OFF-BI-10002867	110	120	2	
9971	9972	2023-07-27	Standard Class	Home Office	United States	Smyrna	Georgia	30080	South	Office Supplies	Art	OFF-AR-10000538	110	140	5	
9972	9973	2022-04-07	Standard Class	Consumer	United States	Houston	Texas	77041	Central	Office Supplies	Envelopes	OFF-EN-10000056	90	100	2	
9973	9974	2023-03-07	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Technology	Phones	TEC-PH-10004080	230	270	5	

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	I
9974	9975	2022-11-17	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Office Supplies	Art	OFF-AR-10004752	20	20	7	
9975	9976	2023-12-16	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Office Supplies	Paper	OFF-PA-10000743	10	10	2	
9976	9977	2022-09-11	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Technology	Phones	TEC-PH-10002496	200	250	2	
9977	9978	2023-12-03	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Office Supplies	Fasteners	OFF-FA-10003467	10	10	7	
9978	9979	2023-06-05	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Office Supplies	Binders	OFF-BI-10003727	10	10	4	
9979	9980	2023-10-12	Standard Class	Home Office	United States	Los Angeles	California	90032	West	Office Supplies	Binders	OFF-BI-10002026	410	440	14	
9980	9981	2023-06-26	Second Class	Consumer	United States	Lafayette	Louisiana	70506	South	Furniture	Tables	FUR-TA-10001039	80	90	1	
9981	9982	2022-11-14	First Class	Consumer	United States	Fairfield	Ohio	45014	East	Office Supplies	Labels	OFF-LA-10004484	20	20	5	
9982	9983	2023-05-27	Standard Class	Consumer	United States	Grand Rapids	Michigan	49505	Central	Office Supplies	Paper	OFF-PA-10002195	40	40	7	
9983	9984	2022-09-14	Standard Class	Consumer	United States	Grand Rapids	Michigan	49505	Central	Technology	Phones	TEC-PH-10001305	80	100	2	
9984	9985	2023-08-04	Standard Class	Consumer	United States	Long Beach	New York	11561	East	Office Supplies	Labels	OFF-LA-10003766	30	30	10	
9985	9986	2023-09-30	Standard Class	Consumer	United States	Long Beach	New York	11561	East	Office Supplies	Supplies	OFF-SU-10000898	50	60	4	
9986	9987	2022-11-30	Standard Class	Consumer	United States	Los Angeles	California	90008	West	Technology	Accessories	TEC-AC-10003399	40	40	1	
9987	9988	2023-04-01	Standard Class	Corporate	United States	Athens	Georgia	30605	South	Technology	Accessories	TEC-AC-10001539	80	80	1	
9988	9989	2022-11-17	Standard Class	Corporate	United States	Athens	Georgia	30605	South	Technology	Phones	TEC-PH-10004006	180	210	5	

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	I
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	FUR-FU-10001889	30	30	3	
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	FUR-FU-10000747	70	90	2	
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	TEC-PH-10003645	220	260	2	
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	OFF-PA-10004041	30	30	4	
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	OFF-AP-10002684	210	240	2	

```
In [86]: # printing the ship mode column unique values as followed from the above
data['Ship Mode'].unique()
```

```
Out[86]: array(['Second Class', 'Standard Class', nan, 'First Class', 'Same Day'],
      dtype=object)
```

```
In [87]: # converting the all cloumns from the data to string to lower
data.columns = data.columns.str.lower()
```

```
In [88]: # replacing the spaces to underscore of the columns headers
data.columns= data.columns.str.replace(' ','_')
data.columns
```

```
Out[88]: Index(['order_id', 'order_date', 'ship_mode', 'segment', 'country', 'city',
      'state', 'postal_code', 'region', 'category', 'sub_category',
      'product_id', 'cost_price', 'list_price', 'quantity',
      'discount_percent'],
      dtype='object')
```

```
In [98]: # printing the above column as followed by
data.head(100)
```

Out[98]:

	Order Id	Order Date	Ship Mode	Segment	Country	City	State	Postal Code	Region	Category	Sub Category	Product Id	cost price	List Price	Quantity	Discount Percent
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240	260	2	2
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600	730	3	3
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10	10	2	!
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780	960	5	2
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20	20	2	!
...
95	96	2023-12-17	Standard Class	Home Office	United States	Portland	Oregon	97206	West	Office Supplies	Binders	OFF-BI-10004738	10	10	1	2
96	97	2022-04-28	Second Class	Home Office	United States	New York City	New York	10009	East	Furniture	Furnishings	FUR-FU-10000629	80	100	7	!
97	98	2023-06-14	First Class	Consumer	United States	San Francisco	California	94122	West	Office Supplies	Binders	OFF-BI-10001721	40	50	3	4
98	99	2023-02-05	Standard Class	Corporate	United States	Saint Paul	Minnesota	55106	Central	Office Supplies	Appliances	OFF-AP-10000358	70	80	6	3
99	100	2023-12-30	Standard Class	Home Office	United States	Chicago	Illinois	60610	Central	Office Supplies	Paper	OFF-PA-10003256	60	60	7	3

100 rows × 16 columns

```
In [29]: # derive new columns discount for the further analysis accordingly
data['discount']=data['list_price']*data['discount_percent']*0.01
data.head(5)
```

Out[29]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	product_id	cost_price	list_p
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240	
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600	
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10	
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780	
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20	

In [30]:

```
# derive new columns sale price for the further analysis accordingly
data['sale_price']= data['list_price']-data['discount_percent']
```

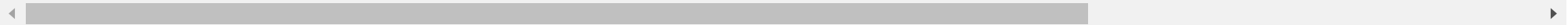
In [37]:

```
# derive new columns profit for the further analysis accordingly
data['profit']= data['sale_price']-data['cost_price']
data
```


Out[37]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	product_id	cost_price
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	240
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	600
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	10
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	780
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	20
...
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	FUR-FU-10001889	30
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	FUR-FU-10000747	70
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	TEC-PH-10003645	220
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	OFF-PA-10004041	30
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	OFF-AP-10002684	210

9994 rows × 19 columns



```
In [49]: #converted order date from object data type to datetime type
data['order_date'] = pd.to_datetime(data['order_date'], format="%Y-%m-%d")
```

```
In [50]: # print the data types for the further analysis
data.dtypes
```

```
Out[50]: order_id          int64
order_date      datetime64[ns]
ship_mode       object
segment         object
country         object
city            object
state           object
postal_code     int64
region          object
category        object
sub_category    object
product_id      object
cost_price      int64
list_price      int64
quantity        int64
discount_percent int64
discount        float64
sale_price      int64
profit          int64
dtype: object
```

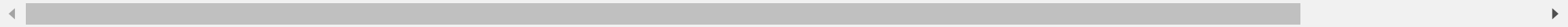
```
In [52]: #drop cost price list price and discount percent columns accordingly
data.drop(columns=['list_price', 'cost_price', 'discount_percent'], inplace=True)
```

```
In [53]: # print the data as followed by the above.
data
```

Out[53]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	product_id	quantity
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	2
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	3
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	2
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	5
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	2
...
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	FUR-FU-10001889	3
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	FUR-FU-10000747	2
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	TEC-PH-10003645	2
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	OFF-PA-10004041	4
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	OFF-AP-10002684	2

9994 rows × 16 columns

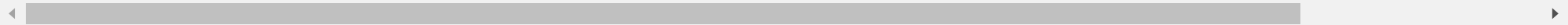


```
In [54]: # printed the final data file
data
```

Out[54]:

	order_id	order_date	ship_mode	segment	country	city	state	postal_code	region	category	sub_category	product_id	quantity
0	1	2023-03-01	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Bookcases	FUR-BO-10001798	2
1	2	2023-08-15	Second Class	Consumer	United States	Henderson	Kentucky	42420	South	Furniture	Chairs	FUR-CH-10000454	3
2	3	2023-01-10	Second Class	Corporate	United States	Los Angeles	California	90036	West	Office Supplies	Labels	OFF-LA-10000240	2
3	4	2022-06-18	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Furniture	Tables	FUR-TA-10000577	5
4	5	2022-07-13	Standard Class	Consumer	United States	Fort Lauderdale	Florida	33311	South	Office Supplies	Storage	OFF-ST-10000760	2
...
9989	9990	2023-02-18	Second Class	Consumer	United States	Miami	Florida	33180	South	Furniture	Furnishings	FUR-FU-10001889	3
9990	9991	2023-03-17	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Furniture	Furnishings	FUR-FU-10000747	2
9991	9992	2022-08-07	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Technology	Phones	TEC-PH-10003645	2
9992	9993	2022-11-19	Standard Class	Consumer	United States	Costa Mesa	California	92627	West	Office Supplies	Paper	OFF-PA-10004041	4
9993	9994	2022-07-17	Second Class	Consumer	United States	Westminster	California	92683	West	Office Supplies	Appliances	OFF-AP-10002684	2

9994 rows × 16 columns



```
In [80]: # exported the clean file by csv format for further analysis in SQL.
data.to_csv(r'D:\Python_sql project\clean_Retaildata.csv', index=False)
```

Conclusion of the EDA:

*Data Analysis Summary & Data Import & Initial Inspection

*1.Imported necessary libraries and read raw CSV files. Displayed the first and last 10 rows of data for an initial overview.

**Data Cleaning

*2.Handled null values using na_values. Standardized column names: converted all text to lowercase and replaced spaces with underscores. Displayed unique values in the 'ship_mode' column. Derived Columns

*3.Created new columns: discount, sale_price, and profit for further analysis.Data Type Conversion

*4.Converted 'order_date' from object type to datetime type.Column Dropping & Export

*5.Dropped unnecessary columns: cost_price, list_price, and discount_percent. Exported the cleaned data to a CSV file for further analysis in SQL.