

## **Project Report**

Introduction To Data Science
Semester — II

# Zudio Sales Analysis

Ву

Aarush C S

Reg No. 2411021240028

Department of Computer Application

Alliance University

Chandapura — Anekal Main Road, Anekal

Bengaluru — 562 106

April 2025

#### Introduction

This project focuses on analyzing and modeling sales data from **Zudio**, a fashion retail brand. The dataset includes transactional and operational records from one or more Zudio outlets.

The objective of the project is to understand key patterns in the sales data and develop a machine learning model to predict or classify important outcomes, such as high or low-performing transactions or product categories.

#### Tasks performed:

- Data Cleaning & Preprocessing
- Exploratory Data Analysis (EDA)
- Machine Learning Model Building & Evaluation
- Visualization for insights

#### **Conclusion / Insights**

- The model trained was a Logistic Regression model.
- Data was successfully cleaned, scaled, and prepared for training.
- The accuracy and F1 scores suggest a decent prediction performance (exact values can be quoted here).
- Confusion matrix helped to visually evaluate model correctness.
- import pandas as pd import numpy as np import matplotlib.pyplot as plt import seaborn as sns from sklearn.model\_selection import train\_test\_split from sklearn.preprocessing import StandardScaler from sklearn.linear\_model import LogisticRegression from sklearn.metrics import confusion\_matrix, classification\_report, accuracy\_score, precision\_score, recall score, f1 score
- df = pd.read\_csv(r"C:\Users\91628\Downloads\Zudio\_sales\_data (1).csv")
- df

•		Store	Country	State	City	Category	Clothing
	Type	\			_		
	0	Zudio	India	Karnataka	Jalna	Women	
	Dresse	es					
	1	Zudio	India	Gujarat	Patna	Kids	
	Skirts	5					
	2	Zudio	India	Uttar Pradesh	Kolkata	Men	
	Hoodi	es					
	3	Zudio	India	West Bengal	Kolkata	Women	
	Dresse	es		_			
	4	Zudio	India	West Bengal	Guwahati	Men	
	Dresse	es		· ·			
	• • •	• • •	• • •	• • •	• • •	• • •	
	• • •						
	7894	Zudio	India	Uttar Pradesh	Chandigarh	Kids	
	Jeans						

	udio	Ind	ia Ut <sup>.</sup>	tar	Pra	desh	. Au	rangat	oad	Kids		
Jackets 7896 Z shirts	udio	Ind	ia	Tar	nil I	Nadu	I	Pι	ıne	Kids		T -
	udio	Ind	ia		D	elhi		Madur	rai	Men		
	'udio	Ind	ia		Guj	arat	В	engalı	ıru	Women		
S	Store	Number	· Pos	tal	Code	e St	ore	Type S	Store	e Open Da	ate	
\												
0		36			11084			wned		2022-06		• • •
1		48		40	<b>3456</b> 2	2		wned		2020-12		• • •
2		97	7	43	3777	3	Re	nted		2023-09	-04	
3		93	3	43	3172	5	0	wned		2024-04	-11	
4		53	L	4:	1537	9	0	wned		2022-05	- 29	
• • •			•			•						
7894		32	2	43	3143	7	0	wned		2024-04	-24	
7895		52			2464			wned		2024-02		
7896		8:			1205			wned		2021-04		• • •
7897		58			3520:			nted		2022-06		
7898		20			2982:			wned		2023-02		• • •
7090		26	9	4,	2902.	5	U	wnea		2023-02	-00	• • •
ID \	ecuri	ity Fea	atures	0r	der :	ID	Orde	r Date	2	Month	Cus	stomer
0 233084			Alarm		882	40	2024	-11-06	5 N	lovember		
1 926472			Alarm		781	45	2024	-06-17	7	June		
2 853255			CCTV		333	63	2024	-08-17	7	August		
3 106146			CCTV		466	82	2024	-03-27	7	March		
4 264070		CCTV,	Alarm		582	42	2024	-04-08	3	April		
• • •			• • •		•			• • •		• • •		
7894		CCTV,	Alarm		390	56	2024	-09-27	7 Se	eptember		
591993 7895 822244			Alarm		854	86	2024	-03-16	5	March		
7896 268376		CCTV,	Alarm		782	90	2024	-12-02	2 0	December		
7897 766283			CCTV		198	83	2024	-07-29	Э	July		
7898 541226		CCTV,	Alarm		914	51	2024	-04-08	3	April		
0 1 2 3 4	Mic Dani St De	stomer chele ( iel Jir teven [ erek On	Ochoa menez Davis Pozco	Pro	oduc <sup>.</sup>	354 741	19 20 14 24	58 78 45	antit	sy Sales 5 1 2 6	39 8 11 59	rofit 916.0 331.2 156.0 925.6
• • •		• •										
7894 7895	Lir	cey And	drews llace				28 18	38 14	- ·	1 2	14	135.2

457 1961

1

784.4

7896

Brandon Baker

```
7897
            John Lewis
                                                          3254.4
                                608 1017
                                                 8
7898 Nicholas Salazar
                                                 5
                                583 2773
                                                          5546.0
[7899 rows x 28 columns]
df.head()
   Store Country
                                      City Category Clothing Type
                          State
\
0
  Zudio
           India
                      Karnataka
                                     Jalna
                                              Women
                                                          Dresses
1 Zudio
           India
                        Gujarat
                                     Patna
                                               Kids
                                                           Skirts
2 Zudio
           India Uttar Pradesh
                                   Kolkata
                                                Men
                                                          Hoodies
3 Zudio
                    West Bengal
           India
                                   Kolkata
                                              Women
                                                          Dresses
4
   Zudio
           India
                    West Bengal Guwahati
                                                Men
                                                          Dresses
   Store Number
                 Postal Code Store Type Store Open Date
0
             30
                      411084
                                   Owned
                                              2022-06-04
1
             48
                      404562
                                   Owned
                                              2020-12-19
2
             97
                      437773
                                  Rented
                                              2023-09-04
3
             93
                      431725
                                   Owned
                                              2024-04-11
4
                      415370
                                              2022-05-29
             51
                                   Owned
                                                           . . .
   Security Features Order ID Order Date
                                               Month Customer ID
\
0
               Alarm
                        88240
                               2024-11-06
                                            November
                                                           233084
                        78145
1
               Alarm
                               2024-06-17
                                                June
                                                          926472
2
                CCTV
                        33363
                               2024-08-17
                                              August
                                                          853255
3
                CCTV
                        46682 2024-03-27
                                               March
                                                          106146
4
         CCTV, Alarm
                        58242
                               2024-04-08
                                               April
                                                          264070
    Customer Name Product ID Price Quantity
                                               Sales Profit
0
    Michele Ochoa
                          354
                               1958
                                            5
                                                     3916.0
  Daniel Jimenez
                          741
                                2078
                                            1
1
                                                      831.2
2
                                            2
     Steven Davis
                          607 1445
                                                     1156.0
3
     Derek Orozco
                          991 2469
                                            6
                                                     5925.6
    Victor Arnold
                          479 1105
                                            6
                                                     2652.0
[5 rows x 28 columns]
df.tail()
      Store Country
                                           City Category Clothing
                              State
Type
      \
7894 Zudio
              India Uttar Pradesh
                                    Chandigarh
                                                    Kids
Jeans
              India Uttar Pradesh Aurangabad
7895 Zudio
                                                    Kids
Jackets
7896 Zudio
              India
                        Tamil Nadu
                                           Pune
                                                    Kids
                                                              T -
shirts
7897 Zudio
                              Delhi
                                        Madurai
              India
                                                     Men
Shoes
7898 Zudio
                           Gujarat
                                      Bengaluru
              India
                                                   Women
Shirts
      Store Number Postal Code Store Type Store Open Date
\
7894
                32
                         431437
                                      Owned
                                                 2024-04-24
7895
                52
                         424643
                                      Owned
                                                 2024-02-03
```

7896

81

412057

Owned

2021-04-11

7898 20 429823 Owned 2023-02-08  Security Features Order ID Order Date Month Custome ID \ 7894 CCTV, Alarm 39056 2024-09-27 September	
ID \	
ID \	
•	ıer
7894 CCTV Alarm 39056 2024-09-27 Sentember	
705+ CCIV, AIGIII 57050 202+ 05 27 September	
591993	
7895 Alarm 85486 2024-03-16 March	
822244	
7896 CCTV, Alarm 78200 2024-12-02 December	
268376	
7897 CCTV 19883 2024-07-29 July	
766283	
7898 CCTV, Alarm 91451 2024-04-08 April	
541226	
Customer Name Product ID Price Quantity Sales Profit	t
7894 Tracey Andrews 262 2838 1 1135.2	
7895 Linda Wallace 537 1814 2 1451.2	2
7896 Brandon Baker 457 1961 1 784.4	4
7897 John Lewis 608 1017 8 3254.4	
7898 Nicholas Salazar 583 2773 5 5546.0	0

[5 rows x 28 columns]

•

#### df.describe()

	ar.describe()			
•	Store Number Staff Count \	Postal Code	Selling Area	Size (sq ft)
	count 7899.000000 7899.000000	7899.000000		7899.000000
	mean 50.666034 11.504747	419982.374984		1112.077478
	std 29.095274 5.194722	11695.237817		193.239371
	min 1.000000 3.000000	400006.000000		800.000000
	25% 25.000000 7.000000	409841.000000		950.000000
	50% 51.000000 12.000000	419988.000000		1100.000000
	75% 76.000000 16.000000	430258.000000		1300.000000
	max 100.000000 20.000000	439985.000000		1400.000000
	Order ID Quantity \	Customer ID	Product ID	Price
	count 7899.000000 7899.000000	7899.000000	7899.000000	7899.000000
	mean 54920.719965 4.519433	549440.714521	602.069376	1761.113432
	std 25894.206850 2.292373	262418.912565	232.066189	716.269447
	min 10008.000000	100153.000000	204.000000	499.000000

```
1.000000
25%
      33192.000000 318516.500000
                                     398.000000
                                                 1136.000000
3.000000
50%
      54873.000000 546012.000000
                                     602.000000
                                                 1782.000000
5.000000
75%
      77285.000000 780304.000000
                                     803.500000
                                                 2373.000000
7.000000
max
      99999.000000 999858.000000
                                     999.000000
                                                 2999.000000
8.000000
```

Sales Profit 7899.000000 count 3170.671376 mean std 2141.249886 min 200.400000 25% 1411.200000 50% 2681.600000 75% 4570.400000 max 9596.800000

- df.info()
- <class 'pandas.core.frame.DataFrame'> RangeIndex: 7899 entries, 0 to 7898
   Data columns (total 28 columns):

# Column		Non-Null Count	Dtype				
0 Store		7899 non-null	object				
1 Country	•	7899 non-null	object				
2 State		7899 non-null	object				
3 City		7899 non-null	object				
4 Categor	ry	7899 non-null	object				
5 Clothir	ig Type	7899 non-null	object				
6 Store N	lumber	7899 non-null	int64				
7 Postal	Code	7899 non-null	int64				
8 Store T	уре	7899 non-null	object				
9 Store C	)pen Date	7899 non-null	object				
10 Selling	g Area Size (sq ft)	7899 non-null	int64				
11 Store M	lanager	7899 non-null	object				
12 Manager	' ID	7899 non-null	object				
13 Store A	ddress	7899 non-null	object				
14 Contact	: Information	7899 non-null	object				
15 Operati	ng Hours	7899 non-null	object				
16 Staff C	Count	7899 non-null	int64				
17 Parking	g Availability	7899 non-null	object				
18 Securit	y Features	5861 non-null	object				
19 Order I	:D	7899 non-null	int64				
20 Order D	ate	7899 non-null	object				
21 Month		7899 non-null	object				
22 Custome	er ID	7899 non-null	int64				
23 Custome	er Name	7899 non-null	object				
24 Product	: ID	7899 non-null	int64				
25 Price		7899 non-null	int64				
26 Quantit	СУ	7899 non-null	int64				
27 Sales P	Profit	7899 non-null	float64				
	` ''	bject(18)					
memory usage: 1.7+ MB							

df.isnull().sum()

	Store Country State City Category Clothing Type Store Number Postal Code Store Type Store Open Date Selling Area Size (sq ft) Store Manager Manager ID Store Address Contact Information Operating Hours Staff Count Parking Availability Security Features Order ID Order Date Month Customer ID Customer Name Product ID Price Quantity Sales Profit	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
•	<pre>dtype: int64  df.nunique()</pre>	Ū
•	Store Country State City Category Clothing Type Store Number Postal Code Store Type Store Open Date Selling Area Size (sq ft) Store Manager Manager ID Store Address Contact Information Operating Hours Staff Count Parking Availability Security Features Order ID Order Date Month Customer ID Customer Name Product ID Price Ouantity	1 8 42 3 11 100 7131 2 1774 7 7476 100 7899 7899 4 18 2 3 7559 338 12 7870 7499 796 2404 8

Quantity

8

Sales Profit 5458

dtype: int64

• df

•	_		Country		Stat	e Ci	ty Category	Clothing
	Type 0	\ Zudio	India	Ka	arnatak	a Jal	na Women	
	Dresse	Zudio	India		Gujara	t Pat	na Kids	
	Skirts	Zudio	India	Uttar	Prades	h Kolka	ta Men	
	Hoodie	Zudio	India	West	t Benga	l Kolka	ta Women	
	Dresse	Zudio	India	West	t Benga	l Guwaha	ti Men	
	Dresse	•••						
	7894	Zudio	India	Uttar	Prades	h Chandiga	rh Kids	
	Jeans 7895 Jacket	Zudio	India	Uttar	Prades	h Aurangab	ad Kids	
	7896	Zudio	India	Tar	mil Nad	u Pu	ne Kids	T -
	shirts 7897 Shoes	Zudio	India		Delh	i Madur	ai Men	
		Zudio	India		Gujara	t Bengalu	ıru Women	
	\	Store	Number	Postal	Code S	tore Type S	tore Open D	ate
	0		30	4.	11084	Owned	2022-06	-04
	1		48		94562	Owned	2020-12	
	2		97		37773	Rented	2023-09	
	3		93		31725	Owned	2024-04	
	4		51		15370	Owned	2022-05	
	• • •		• • •		• • •	• • •		• • • • • •
	7894		32		31437	Owned	2024-04	
	7895		52		24643	Owned	2024-02	
	7896		81	4:	12057	Owned	2021-04	
	7897		58		95203	Rented	2022-06	
	7898		20	42	29823	Owned	2023-02	-08
	ID \	Securi	ity Feat	ures Ord	der ID	Order Date	Month	Customer
	0 233084	ļ.	А	larm	88240	2024-11-06	November	
	1 926472		Α	larm	78145	2024-06-17	' June	
	2 853255		ı	CCTV	33363	2024-08-17	' August	
	3 106146			CCTV	46682	2024-03-27	March	
	4 264076		CCTV, A	larm	58242	2024-04-08	B April	
	• • •			• • •	• • •	• • •	• • •	
	7894		CCTV, A	larm	39056	2024-09-27	' September	

591993	}					
7895	Ala	rm 85486	5 2024-0	3-16	March	
822244	ļ					
7896	CCTV, Ala	rm 78200	0 2024-1	2-02 De	cember	
268376						
7897	CC	TV 19883	3 2024-0	7-29	July	
766283						
7898	CCTV, Ala	rm 91451	L 2024-0	4-08	April	
541226						
			:		6.1	
	Customer Nam			-		
0	Michele Ocho	a 3	354 1958	5	3	3916.0
1	Daniel Jimene	z 7	741 2078	1		831.2
2	Steven Davi	s 6	507 1445	2	1	L156.0
3	Derek Orozc	o 9	991 2469	6	5	925.6
4	Victor Arnol	d 4	479 <b>110</b> 5	6	2	2652.0
	• •	•				
7894	Tracey Andrew	s 2	262 2838	1	1	L135.2
7895	Linda Wallac	e 5	537 1814	2	1	L451.2
7896	Brandon Bake	r 4	457 <b>1</b> 961	1		784.4
7897	John Lewi	s 6	508 1017	8	3	3254.4
7898	Nicholas Salaza	r 5	583 2773	5	5	546.0
Γ <b>7</b> 899	rows v 28 colum	ns 1				

[7899 rows x 28 columns]

 df[['State','City','Category','Clothing Type','Order ID','Customer ID','Product ID','Price','Quantity','Sales Profit']]

•	Customer ID	State	City	Category	Clothing Type	Order ID
	0 Kar	`nataka	Jalna	Women	Dresses	88240
		Gujarat	Patna	Kids	Skirts	78145
	926472 2 Uttar F 853255	Pradesh	Kolkata	Men	Hoodies	33363
		Bengal	Kolkata	Women	Dresses	46682
		Bengal	Guwahati	Men	Dresses	58242
	•••	• • •	• • •	• • •	• • •	• • •
	7894 Uttar F 591993	Pradesh	Chandigarh	Kids	Jeans	39056
	7895 Uttar F 822244	radesh	Aurangabad	Kids	Jackets	85486
	7896 Tami	il Nadu	Pune	Kids	T-shirts	78200
	268376 7897	Delhi	Madurai	Men	Shoes	19883
	766283 7898 0 541226	Gujarat	Bengaluru	Women	Shirts	91451
	Product	t ID Pri	ice Quantit	y Sales	Profit	

	Product ID	Price	Quantity	Sales Profit
0	354	1958	5	3916.0
1	741	2078	1	831.2
2	607	1445	2	1156.0

```
3
            991
                 2469
                            6
                                      5925.6
4
            479
                 1105
                            6
                                      2652.0
                            ...
            . . .
. . .
                  . . .
                                        . . .
7894
                                      1135.2
            262
                 2838
                            1
                             2
7895
            537
                 1814
                                      1451.2
7896
            457
                 1961
                            1
                                      784.4
7897
            608
                 1017
                            8
                                      3254.4
7898
            583
                            5
                 2773
                                      5546.0
```

[7899 rows x 10 columns]

• df[['Order ID','Customer ID','State','City','Category','Clothing Type','Product ID','Price','Quantity','Sales Profit']]

•	0rd	der ID	Customer ID	State	City	Category
	Clothing	Type	\			
	0	88240	233084	Karnataka	Jalna	Women
	Dresses					
	1	78145	926472	Gujarat	Patna	Kids
	Skirts					
	2	33363	853255	Uttar Pradesh	Kolkata	Men
	Hoodies	46600	106146	Uset Bassel	W - 11 - 4 -	Managan
	3	46682	106146	West Bengal	Kolkata	Women
	Dresses	E0242	264070	Wost Bongol	Guwahati	Man
	4 Dresses	58242	264070	West Bengal	Guwanati	Men
	• • •	•••	• • •	•••	• • •	• • •
	7894	39056	591993	Uttar Pradesh	Chandigarh	Kids
	Jeans	33030	33233	occaaacs	chanazga. n	REGS
	7895	85486	822244	Uttar Pradesh	Aurangabad	Kids
	Jackets				Ü	
	7896	78200	268376	Tamil Nadu	Pune	Kids
	T-shirts					
	7897	19883	766283	Delhi	Madurai	Men
	Shoes					
	7898	91451	541226	Gujarat	Bengaluru	Women
	Shirts					

	Product ID	Price	Quantity	Sales Profit
0	354	1958	5	3916.0
1	741	2078	1	831.2
2	607	1445	2	1156.0
3	991	2469	6	5925.6
4	479	1105	6	2652.0
• • •	• • •	• • •	• • •	• • •
7894	262	2838	1	1135.2
7895	537	1814	2	1451.2
7896	457	1961	1	784.4
7897	608	1017	8	3254.4
7898	583	2773	5	5546.0

[7899 rows x 10 columns]

df

<sup>•</sup> df['Discounted price']=df['Price']\*0.7

•	Typo		Country	Sta	te	City	Category	Clothing
	Type 0	Zudio	India	Karnata	ka J	alna	Women	
	Dresse 1	Zudio	India	Gujar	at P	atna	Kids	
	Skirts 2		India	Uttar Prade	sh Kol	kata	Men	
	Hoodie		India	West Beng	al Kol	kata	Women	
	Dresse			•				
	Dresse		IIIUIA	West Beng	ai Guwa	IIIatı	Men	
	• • •	• • •	•••	•	• •	• • •	• • •	
	7894 Jeans	Zudio	India	Uttar Prade	sh Chandi	garh	Kids	
		Zudio	India	Uttar Prade	sh Aurang	abad	Kids	
	7896	Zudio	India	Tamil Na	du	Pune	Kids	T -
		Zudio	India	Del	hi Mad	lurai	Men	
	Shoes 7898 Shirts	Zudio	India	Gujar	at Benga	luru	Women	
	Overden		Number	Postal Code	Store Type	Stor	e Open Da	te
	0	ID \	30	411084	Owned		2022-06-	04
	88240 1		48	404562	Owned		2020-12-	19
	78145 2		97	437773	Rented		2023-09-	04
	33363 3		93	431725	Owned		2024-04-	11
	46682 4		51	415370	Owned		2022-05-	29
	58242		• • •	• • •				
					Ourod		2024-04-	24
	7894 39056		32	431437	Owned			
	7895 85486		52	424643	Owned		2024-02-	03
	7896 78200		81	412057	Owned		2021-04-	11
	7897 19883		58	405203	Rented		2022-06-	17
	7898 91451		20	429823	Owned		2023-02-	08
	D 1	Order		Month Cust	omer ID	Cus	tomer Nam	e
	0	2024-2		\ November	233084	Mic	hele Ocho	a
	354 1	1958 2024-6	<b>06-17</b>	June	926472	Dani	el Jimene	z
	741 2 607	2078 2024-6 1445	08-17	August	853255	St	even Davi	S

3	2024-03-27	March	106146	Derek Orozco
991	2469			
4	2024-04-08	April	264070	Victor Arnold
479	1105			
• • •	• • •	• • •	• • •	•••
 7894	··· 2024-09-27	Sentember	591993	Tracey Andrews
262	2838	September	331333	Truccy Andrews
7895		March	822244	Linda Wallace
537				
7896	2024-12-02	December	268376	Brandon Baker
457	1961			
7897	2024-07-29	July	766283	John Lewis
608				
7898		April	541226	Nicholas Salazar
583	2773			
	Ouantity Sal	os Dnofit	Discounted n	nico
0	Qualitity Sai	3916.0	Discounted p	70.6
0 1	1	831.2		54.6
2	2	1156.0		11.5
3	6	5925.6		28.3
4	6	2652.0		73.5
			,	
 7894	 1	1135.2	19	86.6
7895	2	1451.2		69.8
7896	1	784.4		72.7
7897	8	3254.4		11.9
7898	5	5546.0		41.1
F				

[7899 rows x 29 columns]

• df.drop(columns='Discounted price',inplace=True)

•

df

•	Tuno		Country	State	City	Category	Clothing
	Type 0	\ Zudio	India	Karnataka	Jalna	Women	
	Dresso	es Zudio	India	Gujarat	Patna	Kids	
	Skirt	_		•			
	2 Hoodi	Zudio	India	Uttar Pradesh	Kolkata	Men	
	3	Zudio	India	West Bengal	Kolkata	Women	
	Dresso	es Zudio	India	West Bengal	Guwahati	Men	
	Dress		Inara	Mese bengar	dawanaci	rien	
	• • •	• • •	• • •	• • •	•••	• • •	
	7894	Zudio	India	Uttar Pradesh	Chandigarh	Kids	
	Jeans	Zudio	India	Uttar Pradesh	Aurangahad	Kids	
	Jacke		India	occar Traacsii	Adi diigabad	RIGS	
	7896 shirts	Zudio	India	Tamil Nadu	Pune	Kids	T -
		Zudio	India	Delhi	Madurai	Men	

Shoes 7898 Zudio India Gujarat Bengaluru Women Shirts

Store Number Postal Code Store Type Store Open Date  0 30 411084 Owned 2022-06-04  1 48 404562 Owned 2020-12-19  2 97 437773 Rented 2023-09-04  3 93 431725 Owned 2024-04-11  4 51 415370 Owned 2022-05-29  7894 32 431437 Owned 2024-04-24  7895 52 424643 Owned 2024-02-03  7896 81 412057 Owned 2021-04-11
0       30       411084       Owned       2022-06-04          1       48       404562       Owned       2020-12-19          2       97       437773       Rented       2023-09-04          3       93       431725       Owned       2024-04-11          4       51       415370       Owned       2022-05-29          7894       32       431437       Owned       2024-04-24          7895       52       424643       Owned       2024-02-03
1       48       404562       Owned       2020-12-19          2       97       437773       Rented       2023-09-04          3       93       431725       Owned       2024-04-11          4       51       415370       Owned       2022-05-29          7894       32       431437       Owned       2024-04-24          7895       52       424643       Owned       2024-02-03
2       97       437773       Rented       2023-09-04          3       93       431725       Owned       2024-04-11          4       51       415370       Owned       2022-05-29                  7894       32       431437       Owned       2024-04-24          7895       52       424643       Owned       2024-02-03
3 93 431725 Owned 2024-04-11 4 51 415370 Owned 2022-05-29  7894 32 431437 Owned 2024-04-24 7895 52 424643 Owned 2024-02-03
4 51 415370 Owned 2022-05-29 7894 32 431437 Owned 2024-04-24 7895 52 424643 Owned 2024-02-03
7894 32 431437 Owned 2024-04-24 7895 52 424643 Owned 2024-02-03
7895 52 424643 Owned 2024-02-03
7897 58 405203 Rented 2022-06-17
7898 20 429823 Owned 2023-02-08
Security Features Order ID Order Date Month Custome ID \
0 Alarm 88240 2024-11-06 November 233084
1 Alarm 78145 2024-06-17 June 926472
2 CCTV 33363 2024-08-17 August 853255
3 CCTV 46682 2024-03-27 March 106146
4 CCTV, Alarm 58242 2024-04-08 April
264070
7894 CCTV, Alarm 39056 2024-09-27 September
591993 7895
822244 7896 CCTV, Alarm 78200 2024-12-02 December
268376
7897 CCTV 19883 2024-07-29 July 766283
7898 CCTV, Alarm 91451 2024-04-08 April 541226
Customer Name Product ID Price Quantity Sales Profit
0 Michele Ochoa 354 1958 5 3916.0
1 Daniel Jimenez 741 2078 1 831.2
2 Steven Davis 607 1445 2 1156.0
3 Derek Orozco 991 2469 6 5925.6
4 Victor Arnold 479 1105 6 2652.0
7894 Tracey Andrews 262 2838 1 1135.2
7895 Linda Wallace 537 1814 2 1451.2
7896 Brandon Baker 457 1961 1 784.4
7897 John Lewis 608 1017 8 3254.4
7898 Nicholas Salazar 583 2773 5 5546.0

```
• df.set_index(['Order ID','Category'],inplace=True)
```

• df

****					
	Type \	Store	Country	State	City
Order ID 88240 Dresses	Category Women	Zudio	India	Karnataka	Jalna
78145 Skirts	Kids	Zudio	India	Gujarat	Patna
33363 Hoodies	Men	Zudio	India	Uttar Pradesh	Kolkata
46682 Dresses	Women	Zudio	India	West Bengal	Kolkata
58242 Dresses	Men	Zudio	India	West Bengal	Guwahati
• • •		•••	• • •	•••	•••
39056 Jeans	Kids	Zudio	India	Uttar Pradesh	Chandigarh
85486 Jackets	Kids	Zudio	India	Uttar Pradesh	Aurangabad
78200 T-shirts	Kids	Zudio	India	Tamil Nadu	Pune
19883 Shoes	Men	Zudio	India	Delhi	Madurai
91451 Shirts	Women	Zudio	India	Gujarat	Bengaluru
		Store	Number	Postal Code St	ore Type Store
Open Dat		500.0		. ostar code se	io. c Type Score
Order ID 88240	Category Women		30	411084	Owned
2022-06-					
78145 2020-12-	Kids 19		48	404562	Owned
33363	Men		97	437773	Rented
2023-09- 46682			93	431725	Owned
2024-04-	11				
58242 2022-05-			51	415370	Owned
• • •			• • •	•••	•••
 39056 2024-04-			32	431437	Owned
85486	Kids		52	424643	Owned
2024-02- 78200	Kids		81	412057	Owned
2021-04- 19883	Men		58	405203	Rented
2022-06- 91451 2023-02-	Women		20	429823	Owned
Availabi	lity \	Selli	ng Area :	Size (sq ft) .	Parking

...

Order ID Category

88240 Available	Women			950	• • •	Not
78145 Available	Kids			1200	• • •	Not
33363	Men			1100	• • •	Not
Available 46682	e Women			1400		Not
Available 58242	e Men			1400		Not
Availabl	e					
• • •				•••	•••	
39056 Availabl	Kids e			1200	• • •	Not
85486 Availabl	Kids			950	• • •	Not
78200	Kids			800	• • •	Not
Availablo 19883	e Men			1400		Not
Available 91451	e Women			950		Not
Available				330	• • •	NOC
Customer	TD \	Security Feat	tures	Order Dat	:e	Month
Order ID	Category	,		2024 11 0	No.	
88240 233084	Women	Į.	Alarm	2024-11-0	ob Nov	/ember
78145 926472	Kids	ļ	Alarm	2024-06-1	.7	June
33363 853255	Men		CCTV	2024-08-1	.7	August
46682	Women		CCTV	2024-03-2	27	March
106146 58242	Men	CCTV, A	Alarm	2024-04-0	8	April
264070						
	W 2 4 -	CCTV		2024 00 2		
39056 591993	Kids	CCIV, A	Alarm	2024-09-2	:/ Sep	cember
85486 822244	Kids	ļ	Alarm	2024-03-1	.6	March
78200	Kids	CCTV, A	Alarm	2024-12-0	2 Dec	cember
268376 19883	Men		CCTV	2024-07-2	19	July
766283 91451	Women	CCTV, A	Alarm	2024-04-0	8	April
541226						·
		Customer	Name	Product ID	Price	Quantity
Sales Pro	ofit Category					
88240	Women	Michele (	Ochoa	354	1958	5
3916.0 78145	Kids	Daniel Jin	nenez	741	2078	1
831.2 33363	Men	Steven [	Davis	607	1445	2
1156.0						

46682	Women	Derek Orozco	991	2469	6
5925.6					
58242	Men	Victor Arnold	479	1105	6
2652.0					
• • •		• • •	• • •	• • •	• • •
• • •					
39056	Kids	Tracey Andrews	262	2838	1
1135.2					
85486	Kids	Linda Wallace	537	1814	2
1451.2					
78200	Kids	Brandon Baker	457	1961	1
784.4					
19883	Men	John Lewis	608	1017	8
3254.4					
91451	Women	Nicholas Salazar	583	2773	5
5546.0					

[7899 rows x 26 columns]

- df.reset\_index('Category',inplace=True)
- df

•	Clothing	Category	Store	Country	State	City
	Order ID	туре (				
	88240	Women	Zudio	India	Karnataka	Jalna
	Dresses 78145	Kids	Zudio	India	Gujarat	Patna
	Skirts	KIUS	Zuuio	India	dajarac	raciia
	33363	Men	Zudio	India	Uttar Pradesh	Kolkata
	Hoodies 46682 Dresses	Women	Zudio	India	West Bengal	Kolkata
	58242 Dresses	Men	Zudio	India	West Bengal	Guwahati
	•••		• • •	• • •	•••	• • •
	 39056 Jeans	Kids	Zudio	India	Uttar Pradesh	Chandigarh
	85486	Kids	Zudio	India	Uttar Pradesh	Aurangabad
	Jackets 78200 T-shirts	Kids	Zudio	India	Tamil Nadu	Pune
	19883	Men	Zudio	India	Delhi	Madurai
	Shoes 91451 Shirts	Women	Zudio	India	Gujarat	Bengaluru

78145 48 404562 Owned 2020-12-19 ... 33363 97 437773 Rented 2023-09-04

. . .

46682	9	93	431725	Ov	wned	2024-04-11
 58242	ſ	51	415370	Ou	wned	2022-05-29
•••	•	_	123370	0.		2022 03 23
•••	•	• •	• • •		• • •	•••
39056	3	32	431437	Ov	wned	2024-04-24
85486	<u>.</u>	52	424643	Ov	wned	2024-02-03
78200	8	31	412057	Ov	wned	2021-04-11
 19883		58	405203	Rer	nted	2022-06-17
91451	2	20	429823	Oı	wned	2023-02-08
•••						
Month \	Parking Ava	ailabilit	y Secur	ity Fea	atures	Order Date
Order ID	Not	ا ماداد ا			<b>4.1</b> a.a.m	2024 11 06
88240 November	NOT	Availabl	.e		Alarm	2024-11-06
78145	Not	Availabl	.e		Alarm	2024-06-17
June 33363	Not	Availabl	.e		CCTV	2024-08-17
August						
46682	Not	Availabl	.e		CCTV	2024-03-27
March 58242	Not	Availabl	.e	CCTV,	Alarm	2024-04-08
April 						• • •
• • •						
39056 September		Availabl	.e	CCTV,	Alarm	2024-09-27
85486		Availabl	.e		Alarm	2024-03-16
March 78200	Not	Availabl	.e	CCTV,	Alarm	2024-12-02
December 19883	Not	Availabl	e		CCTV	2024-07-29
July						
91451 April	Not	Availabl	.e	CCTV,	Alarm	2024-04-08
r						
Quantity	Customer ID Sales Profit		omer Na	me Pro	oduct II	) Price
Order ID 88240	233084	Mich	nele Och	oa	354	1 1958
5 3	916.0			<b>.</b>		
78145 1	926472 831.2	Danie	el Jimen	ez	74:	1 2078
33363	853255	Ste	even Dav	is	607	7 1445
2 1 46682	.156.0 106146	Der	ek Oroz	со	99:	1 2469
	925.6	\12 -1	- O.D. A	14	A 7	1105
58242 6 2	264070 652.0	Vict	or Arno	τα	479	9 1105
•••	•••		•	• •	• •	

. . .

. . .

39056	591993	Tracey Andrews	262	2838
1	1135.2			
85486	822244	Linda Wallace	537	1814
2	1451.2			
78200	268376	Brandon Baker	457	1961
1	784.4			
19883	766283	John Lewis	608	1017
8	3254.4			
91451	541226	Nicholas Salazar	583	2773
5	5546.0			

[7899 rows x 27 columns]

- df.loc[df['Quantity']>3,'Quantity']=6

•	Clothing	Category Type \	Store	Country	State	City
	Order ID					
	88240 Dresses	Women	Zudio	India	Karnataka	Jalna
	78145	Kids	Zudio	India	Gujarat	Patna
	Skirts 33363 Hoodies	Men	Zudio	India	Uttar Pradesh	Kolkata
	46682 Dresses	Women	Zudio	India	West Bengal	Kolkata
	58242 Dresses	Men	Zudio	India	West Bengal	Guwahati
	•••	•••	• • •	• • •	•••	•••
	 39056 Jeans	Kids	Zudio	India	Uttar Pradesh	Chandigarh
	85486	Kids	Zudio	India	Uttar Pradesh	Aurangabad
	Jackets 78200 T-shirts	Kids	Zudio	India	Tamil Nadu	Pune
	19883	Men	Zudio	India	Delhi	Madurai
	Shoes 91451 Shirts	Women	Zudio	India	Gujarat	Bengaluru

	Store Number	Postal Code	Store Type	Store Open Date
\ Order ID				
88240	30	411084	Owned	2022-06-04
78145	48	404562	Owned	2020-12-19
33363	97	437773	Rented	2023-09-04
46682	93	431725	Owned	2024-04-11
58242	51	415370	Owned	2022-05-29
• • •				
• • •	• • •	• • •	• • •	• • •
• • •				

39056		32	431437	Oı	wned	2024-04-24		
 85486	!	52	424643	Oı	wned	2024-02-03		
78200	;	81	412057	Oı	wned	2021-04-11		
 19883	!	58	405203	Rei	nted	2022-06-17		
 91451	:	20	429823	Oı	wned	2023-02-08		
• • •								
	Parking Av	ailabilit	y Secur	ity Fe	atures	Order Date		
	\							
Order I		A 2 7 a la 7	_		A 7	2024 11 06		
88240 Novembe		Availabl	.e		Alarm	2024-11-06		
78145		Availabl	e		Alarm	2024-06-17		
June								
33363	Not	Availabl	.e		CCTV	2024-08-17		
August								
46682	Not	Availabl	.e		CCTV	2024-03-27		
March 58242	Not	Availabl	Δ	CCTV	Alarm	2024-04-08		
April	NOC	AVAITADI		cciv,	AIAIIII	2024-04-00		
• • •					• • •	• • •		
• • •								
39056		Availabl	.e	CCTV,	Alarm	2024-09-27		
Septemb		Av. a = 1 a b 1			4.1 a.a.m	2024 02 16		
85486 March	NOL	Availabl	.e		Alarm	2024-03-16		
78200	Not	Availabl	.e	CCTV,	Alarm	2024-12-02		
Decembe	r							
19883	Not	Availabl	e		CCTV	2024-07-29		
July 91451	Not	Availabl	e	CCTV.	Alarm	2024-04-08		
April	1100	71101101		,	, 12G1	202 / 0 / 00		
	Customer ID Customer Name Product ID Price							
Ouantit	y Sales Profi		Ollici IVa		buuct 1	DITIEC		
Order I	•							
	233084	Mich	ele Och	oa	35	4 1958		
	3916.0					4 00=0		
78145 1	926472 831.2	Danie	ı Jimen	ez	74	1 2078		
33363		Ste	ven Dav	is	60	7 1445		
2	1156.0							
46682	106146	Der	ek Oroz	со	99	1 2469		
	5925.6		_					
58242 6	264070 2652.0	Vict	or Arno	Ia	47	9 1105		
•••	2032.0		_	• •				
• • •	•••		•	- <del>-</del>	••			
39056		Trace	y Andre	WS	26	2 2838		
1		1.4	a U-11		F 3:	7 1014		
85486 2	822244 1451.2	Lind	a Matta	ce	53	7 1814		
78200		Bran	don Bak	er	45	7 1961		
1	784.4							

19883	7662	83	John	Lewis	608	1017		
91451	3254.4 5412 5546.0	26 Nich	nolas Sa	lazar	583	2773		
[7899 row	ıs x 27 co	lumns]						
df.iloc[[	df.iloc[[19,1],[1,3]]							
<u></u>	Store	State						
Order ID	Store	State	=					
14350 78145	Zudio R Zudio	-						
df_droppe df_droppe	ed=df.drop ed	na()						
	Category	Store (	Country	9	State	City		
Clothing	Type \		_			•		
Order ID 88240	Women	Zudio	India	Karna	ataka	Jalna		
Dresses					_			
78145 Skirts	Kids	Zudio	India	Gu	jarat	Patna		
33363	Men	Zudio	India	Uttar Pra	adesh	Kolkata		
Hoodies	Llaman	7	Todio	Uo.a≠ Da		Va] ,a+a		
46682 Dresses	Women	Zudio	India	West Be	engai	Kolkata		
58242	Men	Zudio	India	West Be	engal	Guwahati		
Dresses								
• • •	• • •	• • •	• • •		• • •	• • •		
39056	Kids	Zudio	India	Uttar Pra	adesh C	handigarh		
Jeans 85486	Kids	Zudio	India	Uttar Pra	adesh A	urangabad		
Jackets						J		
78200 T-shirts	Kids	Zudio	India	Tamil	Nadu	Pune		
19883	Men	Zudio	India	Г	Delhi	Madurai		
Shoes 91451	Women	Zudio	India	Gu-	ianat	Bengaluru		
Shirts	Wollien	20010	IlluIa	Guj	jarac	bengarui u		
	Stana Nu	mhon Da	ostal Co	do Stono I	Tuna Sta	re Open Dat		
\ Order ID	Store Nu	iliber PC	ostai co	de Store	туре это	re open bac	e	
 88240		30	4110	84 Ov	vned	2022-06-0	4	
 78145		48	4045	62 Ov	vned	2020-12-1	.9	
 33363		97	4377	73 Rer	nted	2023-09-0	4	
• • •							_	

93 431725 Owned 2024-04-11

51 415370 Owned 2022-05-29

. . .

•••

. . .

46682

58242

. . .

39056		32	431437	Oı	wned	2024-04-24		
 85486	!	52	424643	Oı	wned	2024-02-03		
78200	;	81	412057	Oı	wned	2021-04-11		
 19883	!	58	405203	Rei	nted	2022-06-17		
 91451	:	20	429823	Oı	wned	2023-02-08		
• • •								
	Parking Av	ailabilit	y Secur	ity Fe	atures	Order Date		
	\							
Order I		A 2 7 a la 7	_		A 7	2024 11 06		
88240 Novembe		Availabl	.e		Alarm	2024-11-06		
78145		Availabl	e		Alarm	2024-06-17		
June								
33363	Not	Availabl	.e		CCTV	2024-08-17		
August								
46682	Not	Availabl	.e		CCTV	2024-03-27		
March 58242	Not	Availabl	Δ	CCTV	Alarm	2024-04-08		
April	NOC	AVAITADI		cciv,	AIAIIII	2024-04-00		
• • •					• • •	• • •		
• • •								
39056		Availabl	.e	CCTV,	Alarm	2024-09-27		
Septemb		Av. a = 1 a b 1			4.1 a.a.m	2024 02 16		
85486 March	NOL	Availabl	.e		Alarm	2024-03-16		
78200	Not	Availabl	.e	CCTV,	Alarm	2024-12-02		
Decembe	r							
19883	Not	Availabl	e		CCTV	2024-07-29		
July 91451	Not	Availabl	e	CCTV.	Alarm	2024-04-08		
April	1100	71101101		,	, 12G1	202 / 0 / 00		
	Customer ID Customer Name Product ID Price							
Ouantit	y Sales Profi		Ollici IVa		buuct 1	DITIEC		
Order I	•							
	233084	Mich	ele Och	oa	35	4 1958		
	3916.0					4 00=0		
78145 1	926472 831.2	Danie	ı Jimen	ez	74	1 2078		
33363		Ste	ven Dav	is	60	7 1445		
2	1156.0							
46682	106146	Der	ek Oroz	со	99	1 2469		
	5925.6		_					
58242 6	264070 2652.0	Vict	or Arno	Ia	47	9 1105		
•••	2032.0		_	• •				
• • •	•••		•	- <del>-</del>	••			
39056		Trace	y Andre	WS	26	2 2838		
1		1.4	a U-11		F 3:	7 1014		
85486 2	822244 1451.2	Lind	a Matta	ce	53	7 1814		
78200		Bran	don Bak	er	45	7 1961		
1	784.4							

```
19883 766283 John Lewis 608 1017
6 3254.4
91451 541226 Nicholas Salazar 583 2773
6 5546.0

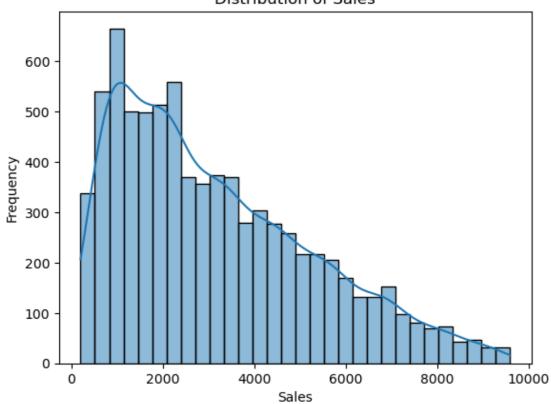
[5861 rows x 27 columns]

import matplotlib.pyplot as plt
import seaborn as sns
```

# Histogram of Sales

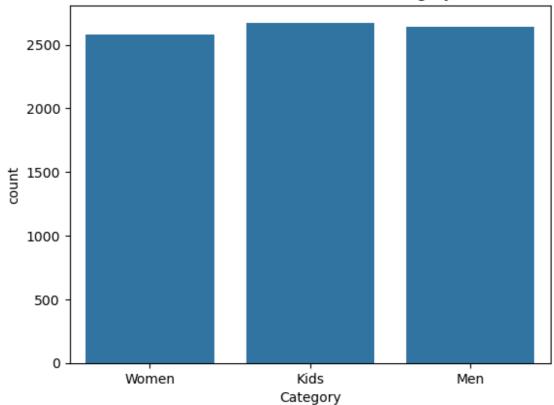
```
sns.histplot(df['Sales Profit'], kde=True)
plt.title("Distribution of Sales")
plt.xlabel("Sales")
plt.ylabel("Frequency")
plt.show()
```

#### Distribution of Sales



sns.countplot(x='Category', data=df)
 plt.title("Count of Products in Each Category")
 plt.show()

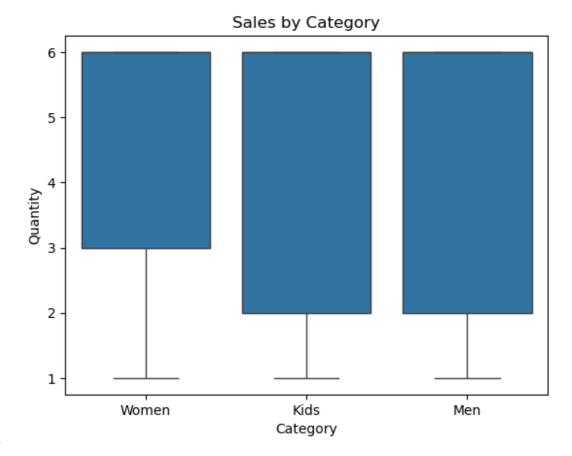
### Count of Products in Each Category



sns.scatterplot(x='Quantity', y='Clothing Type', data=df)
plt.title("Clothing Type sales Quantity")
plt.show()

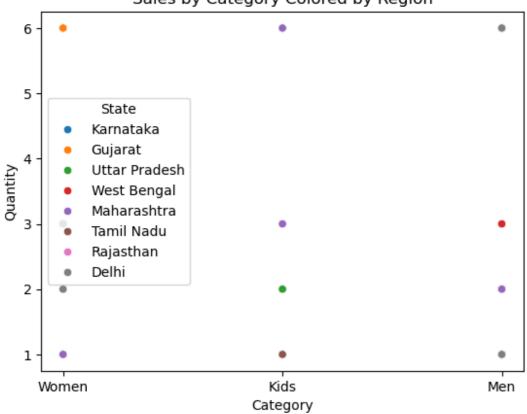


sns.boxplot(x='Category', y='Quantity', data=df)
plt.title("Sales by Category")
plt.show()



sns.scatterplot(x='Category', y='Quantity', hue='State',
data=df)
plt.title("Sales by Category Colored by Region")
plt.show()





•

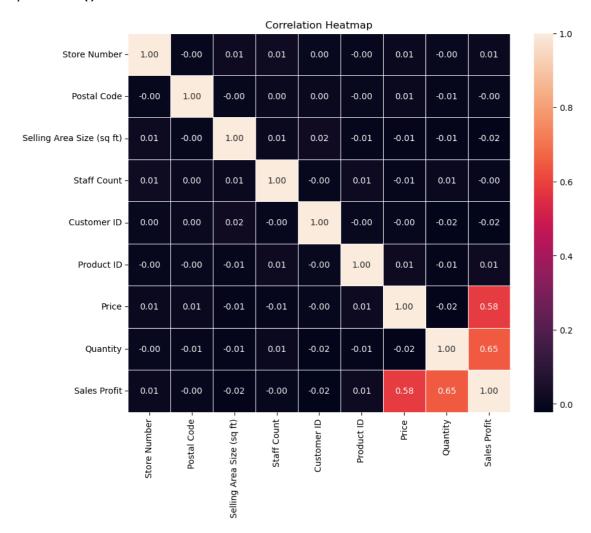
```
import seaborn as sns
import matplotlib.pyplot as plt

# Compute the correlation matrix
corr_matrix = df.corr(numeric_only=True) # only use numeric
columns

# Set the figure size
plt.figure(figsize=(10, 8))

# Create the heatmap
sns.heatmap(corr_matrix, annot=True, linewidths=0.5, fmt=".2f")

# Add a title
plt.title("Correlation Heatmap")
plt.show()
```



X=df[['Price','Quantity']]y=df[['Sales Profit']]

X

•		Price	Quantity
	Order ID		
	88240	1958	6
	78145	2078	1
	33363	1445	2
	46682	2469	6

```
58242
              1105
                           6
   . . .
               . . .
                         . . .
   39056
              2838
                           1
   85486
              1814
                           2
   78200
              1961
                           1
   19883
              1017
                           6
   91451
              2773
                           6
   [7899 rows x 2 columns]
• X_train, X_test, y_train, y_test = train_test_split(X, y,
   test_size=0.2, random_state=42)
 from sklearn.linear model import LinearRegression
   model = LinearRegression()
   model.fit(X_train, y_train)
• LinearRegression()
y_pred = model.predict(X_test)
 from sklearn.metrics import mean_squared_error,
   mean_absolute_error, r2_score
  mae=mean_absolute_error(y_test,y_pred)
   mse=mean_squared_error(y_test,y_pred)
   rmse=np.sqrt(mse)
   r2=r2_score(y_test,y_pred)
   print('mae',mae)
   print('mse',mse)
   print('rmse',rmse)
   print('r2',r2)
  mae 808.5041425003592
   mse 997028.3628041632
   rmse 998.5130759304874
   r2 0.7824154686278934
 y_pred1 = model.predict(X)
 df['y_pred']=y_pred1
  df
            Category Store Country
                                             State
                                                          City
   Clothing Type \
   Order ID
   88240
               Women Zudio
                              India
                                         Karnataka
                                                         Jalna
   Dresses
                Kids Zudio
   78145
                              India
                                           Gujarat
                                                         Patna
   Skirts
                 Men Zudio
                              India Uttar Pradesh
   33363
                                                       Kolkata
   Hoodies
   46682
               Women Zudio
                              India
                                       West Bengal
                                                       Kolkata
   Dresses
   58242
                 Men Zudio
                              India
                                       West Bengal
                                                      Guwahati
```

Dresses ...

39056

Kids Zudio

India Uttar Pradesh Chandigarh

Jeans					
85486 Jackets 78200 T-shirts	Kids	Zudio	India Ut	tar Pradesh	Aurangabad
	Kids	Zudio	India	Tamil Nadu	Pune
19883 Shoes	Men	Zudio	India	Delhi	Madurai
91451 Shirts	Women	Zudio	India	Gujarat	Bengaluru
,	Store Nu	mber Po	stal Code :	Store Type S	Store Open Date
Order ID					
88240		30	411084	Owned	2022-06-04
78145		48	404562	Owned	2020-12-19
33363		97	437773	Rented	2023-09-04
 46682		93	431725	Owned	2024-04-11
 58242		51	415370	Owned	2022-05-29
• • •		• • •		•••	•••
 39056		32	431437	Owned	2024-04-24
 85486		52	424643	Owned	2024-02-03
 78200		81	412057	Owned	2021-04-11
 19883		58	405203	Rented	2022-06-17
 91451		20	429823	Owned	2023-02-08
•••					
\	Security	Feature:	s Order Da	ate Moi	nth Customer ID
Order ID 88240		Aları	n 2024-11	-06 Noveml	222004
78145		Aları			per 233084 une 926472
33363			/ 2024-08	•	
46682 58242	CC	CCT TV, Aları			rch 106146 ril 264070
•••			•	• • •	
39056	CC	TV, Aları Aları	n 2024-09	•	
85486 78200	CC		n 2024-03 n 2024-12		
19883		-	V 2024-07		uly 766283
91451	CC	TV, Aları	n 2024-04	-08 Apı	ril 541226
Profit \		mer Name	Product I	O Price Qua	antity Sales
Order ID 88240	Miche	le Ochoa	354	4 1958	6
3916.0 78145	Daniel	Jimenez	74:	1 2078	1

831.2 33363 1156.0	Steven Davis	607	1445	2
46682 5925.6	Derek Orozco	991	2469	6
58242 2652.0	Victor Arnold	479	1105	6
• • •	•••	• • •	• • •	• • •
 39056 1135.2	Tracey Andrews	262	2838	1
85486	Linda Wallace	537	1814	2
1451.2 78200 784.4	Brandon Baker	457	1961	1
19883 3254.4	John Lewis	608	1017	6
91451 5546.0	Nicholas Salazar	583	2773	6
	y_pred			
Order ID	- <del>-</del>			
88240	4572.576201			
78145	1241.858072			
33363	821.779870			
46682	5484.003712			
58242	3051.152196			
 39056	 2597.405837			
85486	1479.933982			
78200	1033.175061			
19883	2894.194034			
91451	6026.222818			

[7899 rows x 28 columns]