

grocery-sales-forecasting

October 5, 2023

1 Machine Learning Techniques for Sales Forecasting

1.1 Importing Libraries

```
[ ]: %pip install xgboost  
%pip install statsmodels  
%pip install pandas numpy statsmodels  
  
Requirement already satisfied: xgboost in c:\users\srume\anaconda3\lib\site-  
packages (2.0.0)  
Requirement already satisfied: numpy in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from xgboost)  
(1.25.1)  
Requirement already satisfied: scipy in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from xgboost)  
(1.11.1)  
Note: you may need to restart the kernel to use updated packages.  
Requirement already satisfied: statsmodels in c:\users\srume\anaconda3\lib\site-  
packages (0.14.0)  
Requirement already satisfied: numpy>=1.18 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)  
(1.25.1)  
Requirement already satisfied: scipy!=1.9.2,>=1.4 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)  
(1.11.1)  
Requirement already satisfied: pandas>=1.0 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)  
(2.0.3)  
Requirement already satisfied: patsy>=0.5.2 in  
c:\users\srume\anaconda3\lib\site-packages (from statsmodels) (0.5.3)  
Requirement already satisfied: packaging>=21.3 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)  
(23.1)  
Requirement already satisfied: python-dateutil>=2.8.2 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from  
pandas>=1.0->statsmodels) (2.8.2)  
Requirement already satisfied: pytz>=2020.1 in  
c:\users\srume\appdata\roaming\python\python311\site-packages (from
```

```
pandas>=1.0->statsmodels) (2023.3)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from
pandas>=1.0->statsmodels) (2023.3)
Requirement already satisfied: six in
c:\users\srume\appdata\roaming\python\python311\site-packages (from
patsy>=0.5.2->statsmodels) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
Requirement already satisfied: pandas in
c:\users\srume\appdata\roaming\python\python311\site-packages (2.0.3)
Requirement already satisfied: numpy in
c:\users\srume\appdata\roaming\python\python311\site-packages (1.25.1)
Requirement already satisfied: statsmodels in c:\users\srume\anaconda3\lib\site-
packages (0.14.0)
Requirement already satisfied: python-dateutil>=2.8.2 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from pandas)
(2.8.2)
Requirement already satisfied: pytz>=2020.1 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from pandas)
(2023.3)
Requirement already satisfied: tzdata>=2022.1 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from pandas)
(2023.3)
Requirement already satisfied: scipy!=1.9.2,>=1.4 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)
(1.11.1)
Requirement already satisfied: patsy>=0.5.2 in
c:\users\srume\anaconda3\lib\site-packages (from statsmodels) (0.5.3)
Requirement already satisfied: packaging>=21.3 in
c:\users\srume\appdata\roaming\python\python311\site-packages (from statsmodels)
(23.1)
Requirement already satisfied: six in
c:\users\srume\appdata\roaming\python\python311\site-packages (from
patsy>=0.5.2->statsmodels) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

```
[ ]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import scipy.stats as stats
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeRegressor
```

```

from sklearn.tree import plot_tree
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import SGDRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.linear_model import Ridge
from xgboost import XGBRegressor
from sklearn.linear_model import Lasso
from statsmodels.tsa.arima.model import ARIMA
from sklearn.linear_model import BayesianRidge
from sklearn.neighbors import KNeighborsRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
import panel as pn
pn.extension()
import hvplot.pandas
from statsmodels.tsa.stattools import adfuller

```

1.2 Importing Datasets & Read all csv files

1. item_categories.csv - item_category_name, item_category_id
2. items.csv - item_name, item_id, category_id
3. sales_train.csv - date, date_block_num, shop_id, item_id, item_price, item_cnt_day
4. shops.csv - shop_name, shop_id
5. test.csv - ID, shop_id, item_id

```
[ ]: #importing data
item_categories = pd.read_csv('./data-set/item_categories.csv')
items = pd.read_csv('./data-set/items.csv')
sales_train = pd.read_csv('./data-set/sales_train.csv')
shops = pd.read_csv('./data-set/shops.csv')
test = pd.read_csv('./data-set/test.csv')
```

```
[ ]: #checking the shape of the data
print("Shape of item_categories:", item_categories.shape)
print("Shape of items:", items.shape)
print("Shape of sales_train:", sales_train.shape)
print("Shape of shops:", shops.shape)
print("Shape of test:", test.shape)
```

Shape of item_categories: (84, 2)
 Shape of items: (22170, 3)
 Shape of sales_train: (2935849, 6)

```
Shape of shops: (60, 2)
Shape of test: (214200, 3)
```

```
[ ]: #checking the columns of the data
print("\n\nColumns of item_categories:\n")
print(item_categories.info())

print("-----")

print("\n\nColumns of items:\n")
print(items.info())

print("-----")

print("\n\nColumns of sales_train:\n")
print(sales_train.info())

print("-----")

print("\n\nColumns of shops:\n")
print(shops.info())

print("-----")

print("\n\nColumns of test:\n")
print(test.info())
```

```
Columns of item_categories:
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 84 entries, 0 to 83
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  --  
 0   item_category_name  84 non-null    object  
 1   item_category_id    84 non-null    int64  
dtypes: int64(1), object(1)
memory usage: 1.4+ KB
None
-----
```

```
Columns of items:
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22170 entries, 0 to 22169
```

```
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          --    
 0   item_name    22170 non-null   object 
 1   item_id     22170 non-null   int64  
 2   category_id 22170 non-null   int64  
dtypes: int64(2), object(1)
memory usage: 519.7+ KB
None
```

Columns of sales_train:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2935849 entries, 0 to 2935848
Data columns (total 6 columns):
 #   Column      Dtype  
 ---  --          --    
 0   date        object 
 1   date_block_num  int64  
 2   shop_id     int64  
 3   item_id     int64  
 4   item_price   float64
 5   item_cnt_day float64
dtypes: float64(2), int64(3), object(1)
memory usage: 134.4+ MB
None
```

Columns of shops:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 60 entries, 0 to 59
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
 ---  --          -----          --    
 0   shop_name   60 non-null    object 
 1   shop_id     60 non-null    int64  
dtypes: int64(1), object(1)
memory usage: 1.1+ KB
None
```

Columns of test:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 214200 entries, 0 to 214199
Data columns (total 3 columns):
 #   Column   Non-Null Count   Dtype  
---  -- 
 0   ID        214200 non-null    int64  
 1   shop_id   214200 non-null    int64  
 2   item_id   214200 non-null    int64  
dtypes: int64(3)
memory usage: 4.9 MB
None
```

```
[ ]: #checking the head and tail of the data
```

```
print("\n\nHead of item_categories:\n")
print(item_categories.head())

print("\n\nTail of item_categories:\n")
print(item_categories.tail())

print("-----")

print("\n\nHead of items:\n")
print(items.head())

print("\n\nTail of items:\n")
print(items.tail())

print("-----")

print("\n\nHead of sales_train:\n")
print(sales_train.head())

print("\n\nTail of sales_train:\n")
print(sales_train.tail())

print("-----")

print("\n\nHead of shops:\n")
print(shops.head())

print("\n\nTail of shops:\n")
print(shops.tail())

print("-----")

print("\n\nHead of test:\n")
```

```

print(test.head())
print("\n\nTail of test:\n")
print(test.tail())

```

Head of item_categories:

	item_category_name	item_category_id
0	PC - Headsets / Headphones	0
1	Accessories - PS2	1
2	Accessories - PS3	2
3	Accessories - PS4	3
4	Accessories - PSP	4

Tail of item_categories:

	item_category_name	item_category_id
79	Service	79
80	Service - Tickets	80
81	Blank media (spire)	81
82	Blank media (piece)	82
83	Batteries	83

Head of items:

	item_name	item_id	category_id
0	!! IN THE POWER OF HAPPINESS (PLAST) D	0	40
1	! ABBYY FineReader 12 Professional Edition Ful...	1	76
2	*** IN THE GLORY OF THE GLORY (UNV) D	2	40
3	*** BLUE WAVE (Univ) D	3	40
4	*** BOX (GLASS) D	4	40

Tail of items:

	item_name	item_id	category_id
22165	Nuclear Titbit 2 [PC, Digital Version]	22165	31
22166	Query language 1C: Enterprise [Digital version]	22166	54
22167	The query language is 1C: Enterprise 8 (+ CD)...	22167	49
22168	Egg for Little Inu	22168	62
22169	Egg of the Dragon (Game of Thrones)	22169	69

Head of sales_train:

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day
0	02.01.2013	0	59	22154	999.00	1.0
1	03.01.2013	0	25	2552	899.00	1.0
2	05.01.2013	0	25	2552	899.00	-1.0
3	06.01.2013	0	25	2554	1709.05	1.0
4	15.01.2013	0	25	2555	1099.00	1.0

Tail of sales_train:

	date	date_block_num	shop_id	item_id	item_price	\
2935844	10.10.2015	33	25	7409	299.0	
2935845	09.10.2015	33	25	7460	299.0	
2935846	14.10.2015	33	25	7459	349.0	
2935847	22.10.2015	33	25	7440	299.0	
2935848	03.10.2015	33	25	7460	299.0	
				item_cnt_day		
2935844				1.0		
2935845				1.0		
2935846				1.0		
2935847				1.0		
2935848				1.0		

Head of shops:

	shop_name	shop_id
0	! Yakutsk Ordzhonikidze, 56 francs	0
1	! Yakutsk TC "Central" fran	1
2	Adygea TC "Mega"	2
3	Balashikha TC "Oktyabr-Kinomir"	3
4	Volga TC "Volga Mall"	4

Tail of shops:

	shop_name	shop_id
55	Digital warehouse 1C-Online	55
56	Chekhov TC" Karnaval "	56
57	Yakutsk Ordzhonikidze, 56	57
58	Yakutsk TC" Central "	58
59	Yaroslavl TC" Altair "	59

Head of test:

```
ID  shop_id  item_id
0   0         5       5037
1   1         5       5320
2   2         5       5233
3   3         5       5232
4   4         5       5268
```

Tail of test:

```
ID  shop_id  item_id
214195 214195    45    18454
214196 214196    45    16188
214197 214197    45    15757
214198 214198    45    19648
214199 214199    45    969
```

1.3 Data Preprocessing & Feature Engineering

```
[ ]: #merging the data for better understand the data
[ ]: #Merge sales_train.csv with items.csv on the "item_id" column
sales_with_items = sales_train.merge(items, on='item_id', how='left')
print("\n\nHead of sales_with_items:\n")
print(sales_with_items.head(20))
print(sales_with_items.shape)
```

Head of sales_with_items:

```
date  date_block_num  shop_id  item_id  item_price  item_cnt_day \
0    02.01.2013        0       59    22154     999.00      1.0
1    03.01.2013        0       25    2552      899.00      1.0
2    05.01.2013        0       25    2552      899.00     -1.0
3    06.01.2013        0       25    2554     1709.05      1.0
4    15.01.2013        0       25    2555     1099.00      1.0
5    10.01.2013        0       25    2564      349.00      1.0
6    02.01.2013        0       25    2565      549.00      1.0
7    04.01.2013        0       25    2572      239.00      1.0
8    11.01.2013        0       25    2572      299.00      1.0
9    03.01.2013        0       25    2573      299.00      3.0
10   03.01.2013        0       25    2574      399.00      2.0
11   05.01.2013        0       25    2574      399.00      1.0
```

12	07.01.2013	0	25	2574	399.00	1.0
13	08.01.2013	0	25	2574	399.00	2.0
14	10.01.2013	0	25	2574	399.00	1.0
15	11.01.2013	0	25	2574	399.00	2.0
16	13.01.2013	0	25	2574	399.00	1.0
17	16.01.2013	0	25	2574	399.00	1.0
18	26.01.2013	0	25	2574	399.00	1.0
19	27.01.2013	0	25	2574	399.00	1.0

		item_name	category_id
0		SCENE 2012 (BD)	37
1	DEEP PURPLE	The House Of Blue Light LP	58
2	DEEP PURPLE	The House Of Blue Light LP	58
3	DEEP PURPLE	Who Do You Think We Are LP	58
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59
6	DEEP PURPLE	Stormbringer (firms).	56
7		DEFTONES Koi No Yakan	55
8		DEFTONES Koi No Yakan	55
9	DEL REY LANA	Born To Die	55
10	DEL REY LANA	Born To Die The Paradise Editio...	55
11	DEL REY LANA	Born To Die The Paradise Editio...	55
12	DEL REY LANA	Born To Die The Paradise Editio...	55
13	DEL REY LANA	Born To Die The Paradise Editio...	55
14	DEL REY LANA	Born To Die The Paradise Editio...	55
15	DEL REY LANA	Born To Die The Paradise Editio...	55
16	DEL REY LANA	Born To Die The Paradise Editio...	55
17	DEL REY LANA	Born To Die The Paradise Editio...	55
18	DEL REY LANA	Born To Die The Paradise Editio...	55
19	DEL REY LANA	Born To Die The Paradise Editio...	55

(2935849, 8)

```
[ ]: #Merge the result with item_categories.csv on the "category_id"
sales_with_items_and_categories = sales_with_items.merge(item_categories, how='left',
    right_on='item_category_id', left_on='category_id', how='left')
print("\n\nHead of sales_with_items_and_categories:\n")
print(sales_with_items_and_categories.head(20))
print(sales_with_items_and_categories.shape)
```

Head of sales_with_items_and_categories:

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day	\
0	02.01.2013	0	59	22154	999.00	1.0	
1	03.01.2013	0	25	2552	899.00	1.0	
2	05.01.2013	0	25	2552	899.00	-1.0	
3	06.01.2013	0	25	2554	1709.05	1.0	

4	15.01.2013	0	25	2555	1099.00	1.0
5	10.01.2013	0	25	2564	349.00	1.0
6	02.01.2013	0	25	2565	549.00	1.0
7	04.01.2013	0	25	2572	239.00	1.0
8	11.01.2013	0	25	2572	299.00	1.0
9	03.01.2013	0	25	2573	299.00	3.0
10	03.01.2013	0	25	2574	399.00	2.0
11	05.01.2013	0	25	2574	399.00	1.0
12	07.01.2013	0	25	2574	399.00	1.0
13	08.01.2013	0	25	2574	399.00	2.0
14	10.01.2013	0	25	2574	399.00	1.0
15	11.01.2013	0	25	2574	399.00	2.0
16	13.01.2013	0	25	2574	399.00	1.0
17	16.01.2013	0	25	2574	399.00	1.0
18	26.01.2013	0	25	2574	399.00	1.0
19	27.01.2013	0	25	2574	399.00	1.0

		item_name	category_id	\
0		SCENE 2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light LP	58	
2	DEEP PURPLE	The House Of Blue Light LP	58	
3	DEEP PURPLE	Who Do You Think We Are LP	58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE	Stormbringer (firms).	56	
7	DEFTONES	Koi No Yakan	55	
8	DEFTONES	Koi No Yakan	55	
9	DEL REY LANA	Born To Die	55	
10	DEL REY LANA	Born To Die The Paradise Editio...	55	
11	DEL REY LANA	Born To Die The Paradise Editio...	55	
12	DEL REY LANA	Born To Die The Paradise Editio...	55	
13	DEL REY LANA	Born To Die The Paradise Editio...	55	
14	DEL REY LANA	Born To Die The Paradise Editio...	55	
15	DEL REY LANA	Born To Die The Paradise Editio...	55	
16	DEL REY LANA	Born To Die The Paradise Editio...	55	
17	DEL REY LANA	Born To Die The Paradise Editio...	55	
18	DEL REY LANA	Born To Die The Paradise Editio...	55	
19	DEL REY LANA	Born To Die The Paradise Editio...	55	

	item_category_name	item_category_id
0	Cinema - Blu-Ray	37
1	Music - Vinyl	58
2	Music - Vinyl	58
3	Music - Vinyl	58
4	Music - CD of corporate production	56
5	Music - Music video	59
6	Music - CD of corporate production	56
7	Music - CD of local production	55

```

8      Music - CD of local production      55
9      Music - CD of local production      55
10     Music - CD of local production      55
11     Music - CD of local production      55
12     Music - CD of local production      55
13     Music - CD of local production      55
14     Music - CD of local production      55
15     Music - CD of local production      55
16     Music - CD of local production      55
17     Music - CD of local production      55
18     Music - CD of local production      55
19     Music - CD of local production      55
(2935849, 10)

```

```

[ ]: # Check if the two columns are the same
if sales_with_items_and_categories['item_category_id'].
    ~equals(sales_with_items_and_categories['category_id']):
    # If they are the same, drop one of the columns
    sales_with_items_and_categories.drop(columns=['item_category_id'], □
    ~inplace=True)

```

```

[ ]: print("\n\nHead of sales_with_items_and_categories:\n")
print(sales_with_items_and_categories.head(20))
print(sales_with_items_and_categories.shape)

```

Head of sales_with_items_and_categories:

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day	\
0	02.01.2013	0	59	22154	999.00	1.0	
1	03.01.2013	0	25	2552	899.00	1.0	
2	05.01.2013	0	25	2552	899.00	-1.0	
3	06.01.2013	0	25	2554	1709.05	1.0	
4	15.01.2013	0	25	2555	1099.00	1.0	
5	10.01.2013	0	25	2564	349.00	1.0	
6	02.01.2013	0	25	2565	549.00	1.0	
7	04.01.2013	0	25	2572	239.00	1.0	
8	11.01.2013	0	25	2572	299.00	1.0	
9	03.01.2013	0	25	2573	299.00	3.0	
10	03.01.2013	0	25	2574	399.00	2.0	
11	05.01.2013	0	25	2574	399.00	1.0	
12	07.01.2013	0	25	2574	399.00	1.0	
13	08.01.2013	0	25	2574	399.00	2.0	
14	10.01.2013	0	25	2574	399.00	1.0	
15	11.01.2013	0	25	2574	399.00	2.0	
16	13.01.2013	0	25	2574	399.00	1.0	
17	16.01.2013	0	25	2574	399.00	1.0	

18	26.01.2013	0	25	2574	399.00	1.0
19	27.01.2013	0	25	2574	399.00	1.0

			item_name	category_id	\
0			SCENE 2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light	LP	58	
2	DEEP PURPLE	The House Of Blue Light	LP	58	
3	DEEP PURPLE	Who Do You Think We Are	LP	58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).		56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...		59	
6	DEEP PURPLE	Stormbringer (firms).		56	
7	DEFTONES	Koi No Yakan		55	
8	DEFTONES	Koi No Yakan		55	
9	DEL REY LANA	Born To Die		55	
10	DEL REY LANA	Born To Die	The Paradise Editio...	55	
11	DEL REY LANA	Born To Die	The Paradise Editio...	55	
12	DEL REY LANA	Born To Die	The Paradise Editio...	55	
13	DEL REY LANA	Born To Die	The Paradise Editio...	55	
14	DEL REY LANA	Born To Die	The Paradise Editio...	55	
15	DEL REY LANA	Born To Die	The Paradise Editio...	55	
16	DEL REY LANA	Born To Die	The Paradise Editio...	55	
17	DEL REY LANA	Born To Die	The Paradise Editio...	55	
18	DEL REY LANA	Born To Die	The Paradise Editio...	55	
19	DEL REY LANA	Born To Die	The Paradise Editio...	55	

	item_category_name
0	Cinema - Blu-Ray
1	Music - Vinyl
2	Music - Vinyl
3	Music - Vinyl
4	Music - CD of corporate production
5	Music - Music video
6	Music - CD of corporate production
7	Music - CD of local production
8	Music - CD of local production
9	Music - CD of local production
10	Music - CD of local production
11	Music - CD of local production
12	Music - CD of local production
13	Music - CD of local production
14	Music - CD of local production
15	Music - CD of local production
16	Music - CD of local production
17	Music - CD of local production
18	Music - CD of local production
19	Music - CD of local production

(2935849, 9)

```
[ ]: #Merge the result with shops.csv on the "shop_id"
final_dataset = sales_with_items_and_categories.merge(shops, on='shop_id', how='left')
print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	date_block_num	shop_id	item_id	item_price	item_cnt_day	\
0	02.01.2013	0	59	22154	999.00	1.0	
1	03.01.2013	0	25	2552	899.00	1.0	
2	05.01.2013	0	25	2552	899.00	-1.0	
3	06.01.2013	0	25	2554	1709.05	1.0	
4	15.01.2013	0	25	2555	1099.00	1.0	
5	10.01.2013	0	25	2564	349.00	1.0	
6	02.01.2013	0	25	2565	549.00	1.0	
7	04.01.2013	0	25	2572	239.00	1.0	
8	11.01.2013	0	25	2572	299.00	1.0	
9	03.01.2013	0	25	2573	299.00	3.0	
10	03.01.2013	0	25	2574	399.00	2.0	
11	05.01.2013	0	25	2574	399.00	1.0	
12	07.01.2013	0	25	2574	399.00	1.0	
13	08.01.2013	0	25	2574	399.00	2.0	
14	10.01.2013	0	25	2574	399.00	1.0	
15	11.01.2013	0	25	2574	399.00	2.0	
16	13.01.2013	0	25	2574	399.00	1.0	
17	16.01.2013	0	25	2574	399.00	1.0	
18	26.01.2013	0	25	2574	399.00	1.0	
19	27.01.2013	0	25	2574	399.00	1.0	

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
2	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE Perihelion: Live In Concert DVD (C... 59		
6	DEEP PURPLE Stormbringer (firms). 56		
7	DEFTONES Koi No Yakan 55		
8	DEFTONES Koi No Yakan 55		
9	DEL REY LANA Born To Die 55		
10	DEL REY LANA Born To Die The Paradise Editio... 55		
11	DEL REY LANA Born To Die The Paradise Editio... 55		
12	DEL REY LANA Born To Die The Paradise Editio... 55		
13	DEL REY LANA Born To Die The Paradise Editio... 55		

```

14  DEL REY LANA Born To Die The Paradise Editio...      55
15  DEL REY LANA Born To Die The Paradise Editio...      55
16  DEL REY LANA Born To Die The Paradise Editio...      55
17  DEL REY LANA Born To Die The Paradise Editio...      55
18  DEL REY LANA Born To Die The Paradise Editio...      55
19  DEL REY LANA Born To Die The Paradise Editio...      55

```

	item_category_name	shop_name
0	Cinema - Blu-Ray	Yaroslavl TEC" Altair "
1	Music - Vinyl	Moscow TEC" Atrium "
2	Music - Vinyl	Moscow TEC" Atrium "
3	Music - Vinyl	Moscow TEC" Atrium "
4	Music - CD of corporate production	Moscow TEC" Atrium "
5	Music - Music video	Moscow TEC" Atrium "
6	Music - CD of corporate production	Moscow TEC" Atrium "
7	Music - CD of local production	Moscow TEC" Atrium "
8	Music - CD of local production	Moscow TEC" Atrium "
9	Music - CD of local production	Moscow TEC" Atrium "
10	Music - CD of local production	Moscow TEC" Atrium "
11	Music - CD of local production	Moscow TEC" Atrium "
12	Music - CD of local production	Moscow TEC" Atrium "
13	Music - CD of local production	Moscow TEC" Atrium "
14	Music - CD of local production	Moscow TEC" Atrium "
15	Music - CD of local production	Moscow TEC" Atrium "
16	Music - CD of local production	Moscow TEC" Atrium "
17	Music - CD of local production	Moscow TEC" Atrium "
18	Music - CD of local production	Moscow TEC" Atrium "
19	Music - CD of local production	Moscow TEC" Atrium "

(2935849, 10)

```
[ ]: #checks the columns of the final dataset
print("\n\nColumns of final_dataset:\n")
print(final_dataset.info())
```

Columns of final_dataset:

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2935849 entries, 0 to 2935848
Data columns (total 10 columns):
 #   Column            Dtype  
 --- 
 0   date              object 
 1   date_block_num    int64  
 2   shop_id           int64  
 3   item_id           int64  
 4   item_price        float64

```

```
5    item_cnt_day      float64
6    item_name          object
7    category_id        int64
8    item_category_name object
9    shop_name          object
dtypes: float64(2), int64(4), object(4)
memory usage: 224.0+ MB
None
```

```
[ ]: #prints the date and date_block_num column to check whether they are related
columns_to_print = ['date', 'date_block_num']
print(final_dataset[columns_to_print])
```

	date	date_block_num
0	02.01.2013	0
1	03.01.2013	0
2	05.01.2013	0
3	06.01.2013	0
4	15.01.2013	0
...
2935844	10.10.2015	33
2935845	09.10.2015	33
2935846	14.10.2015	33
2935847	22.10.2015	33
2935848	03.10.2015	33

[2935849 rows x 2 columns]

```
[ ]: # Rename the column
final_dataset.rename(columns={'date_block_num': 'month_num'}, inplace=True)
```

```
[ ]: #Rename the column
final_dataset.rename(columns={'item_cnt_day': 'item_cnt_month'}, inplace=True)
```

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00		1.0
1	03.01.2013	0	25	2552	899.00		1.0
2	05.01.2013	0	25	2552	899.00		-1.0
3	06.01.2013	0	25	2554	1709.05		1.0
4	15.01.2013	0	25	2555	1099.00		1.0
5	10.01.2013	0	25	2564	349.00		1.0

6	02.01.2013	0	25	2565	549.00	1.0
7	04.01.2013	0	25	2572	239.00	1.0
8	11.01.2013	0	25	2572	299.00	1.0
9	03.01.2013	0	25	2573	299.00	3.0
10	03.01.2013	0	25	2574	399.00	2.0
11	05.01.2013	0	25	2574	399.00	1.0
12	07.01.2013	0	25	2574	399.00	1.0
13	08.01.2013	0	25	2574	399.00	2.0
14	10.01.2013	0	25	2574	399.00	1.0
15	11.01.2013	0	25	2574	399.00	2.0
16	13.01.2013	0	25	2574	399.00	1.0
17	16.01.2013	0	25	2574	399.00	1.0
18	26.01.2013	0	25	2574	399.00	1.0
19	27.01.2013	0	25	2574	399.00	1.0

		item_name	category_id	\
0		SCENE 2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light LP	58	
2	DEEP PURPLE	The House Of Blue Light LP	58	
3	DEEP PURPLE	Who Do You Think We Are LP	58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE	Stormbringer (firms).	56	
7	DEFTONES	Koi No Yakan	55	
8	DEFTONES	Koi No Yakan	55	
9	DEL REY LANA	Born To Die	55	
10	DEL REY LANA	Born To Die The Paradise Editio...	55	
11	DEL REY LANA	Born To Die The Paradise Editio...	55	
12	DEL REY LANA	Born To Die The Paradise Editio...	55	
13	DEL REY LANA	Born To Die The Paradise Editio...	55	
14	DEL REY LANA	Born To Die The Paradise Editio...	55	
15	DEL REY LANA	Born To Die The Paradise Editio...	55	
16	DEL REY LANA	Born To Die The Paradise Editio...	55	
17	DEL REY LANA	Born To Die The Paradise Editio...	55	
18	DEL REY LANA	Born To Die The Paradise Editio...	55	
19	DEL REY LANA	Born To Die The Paradise Editio...	55	

	item_category_name	shop_name
0	Cinema - Blu-Ray	Yaroslavl TC" Altair "
1	Music - Vinyl	Moscow TEC" Atrium "
2	Music - Vinyl	Moscow TEC" Atrium "
3	Music - Vinyl	Moscow TEC" Atrium "
4	Music - CD of corporate production	Moscow TEC" Atrium "
5	Music - Music video	Moscow TEC" Atrium "
6	Music - CD of corporate production	Moscow TEC" Atrium "
7	Music - CD of local production	Moscow TEC" Atrium "
8	Music - CD of local production	Moscow TEC" Atrium "
9	Music - CD of local production	Moscow TEC" Atrium "

```
10      Music - CD of local production      Moscow TEC" Atrium "
11      Music - CD of local production      Moscow TEC" Atrium "
12      Music - CD of local production      Moscow TEC" Atrium "
13      Music - CD of local production      Moscow TEC" Atrium "
14      Music - CD of local production      Moscow TEC" Atrium "
15      Music - CD of local production      Moscow TEC" Atrium "
16      Music - CD of local production      Moscow TEC" Atrium "
17      Music - CD of local production      Moscow TEC" Atrium "
18      Music - CD of local production      Moscow TEC" Atrium "
19      Music - CD of local production      Moscow TEC" Atrium "
(2935849, 10)
```

```
[ ]: #checks the columns of the final dataset
print("\n\nColumns of final_dataset:\n")
print(final_dataset.info())
```

Columns of final_dataset:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2935849 entries, 0 to 2935848
Data columns (total 10 columns):
 #   Column           Dtype  
--- 
 0   date             object 
 1   month_num        int64  
 2   shop_id          int64  
 3   item_id          int64  
 4   item_price       float64
 5   item_cnt_month   float64
 6   item_name        object 
 7   category_id     int64  
 8   item_category_name  object 
 9   shop_name        object 
dtypes: float64(2), int64(4), object(4)
memory usage: 224.0+ MB
None
```

```
[ ]: #export the final dataset to csv file
final_dataset.to_csv('./data-set/output/final_dataset_without_cleaning.csv', index=False)
```

```
[ ]: #Data Cleaning

#checking for missing values
print("\n\nMissing values in final_dataset:\n")
print(final_dataset.isnull().sum())
```

```
Missing values in final_dataset:
```

```
date          0
month_num     0
shop_id       0
item_id       0
item_price    0
item_cnt_month 0
item_name     0
category_id   0
item_category_name 0
shop_name     0
dtype: int64
```

```
[ ]: #checking for null values
print("\n\nNull values in final_dataset:\n")
print(final_dataset.isnull().sum())
```

```
Null values in final_dataset:
```

```
date          0
month_num     0
shop_id       0
item_id       0
item_price    0
item_cnt_month 0
item_name     0
category_id   0
item_category_name 0
shop_name     0
dtype: int64
```

```
[ ]: print(final_dataset.shape)
```



```
(2935849, 10)
```

```
[ ]: #handles the missing values in final_dataset
final_dataset['item_name'].fillna('Unknown', inplace=True)
final_dataset['item_category_name'].fillna('Unknown', inplace=True)
```

```
[ ]: print(final_dataset.shape)
```



```
(2935849, 10)
```

```
[ ]: #removes duplicates rows in final_dataset
final_dataset.drop_duplicates(inplace=True)

[ ]: print(final_dataset.shape)

(2935843, 10)

[ ]: #checks and solves the data type of the columns
print("\n\nData types of final_dataset:\n")
print(final_dataset.dtypes)
```

Data types of final_dataset:

```
date                object
month_num           int64
shop_id              int64
item_id              int64
item_price           float64
item_cnt_month       float64
item_name             object
category_id          int64
item_category_name   object
shop_name             object
dtype: object
```

```
[ ]: # #seems like item_cnt_month should be int64
final_dataset['item_cnt_month'] = final_dataset['item_cnt_month'].  
    ↪astype('int64')
```

```
[ ]: print(final_dataset.dtypes)
```

```
date                object
month_num           int64
shop_id              int64
item_id              int64
item_price           float64
item_cnt_month       int64
item_name             object
category_id          int64
item_category_name   object
shop_name             object
dtype: object
```

```
[ ]: #prints item_cnt_month column to check whether it is int64
print(final_dataset['item_cnt_month'].head(30))
```

```
0      1
```

```
1      1
2     -1
3      1
4      1
5      1
6      1
7      1
8      1
9      3
10     2
11     1
12     1
13     2
14     1
15     2
16     1
17     1
18     1
19     1
20     1
21     1
22     1
23     1
24     1
25     1
26     1
27     1
28     1
29     1
Name: item_cnt_month, dtype: int64
```

```
[ ]: print(final_dataset.shape)
```

```
(2935843, 10)
```

```
[ ]: #removes -1 and 307980 from item_cnt_month column because it is an outlier
#it is not possible to sell -1 and 307980 items in a day because 307980 is the
#total number of items sold in a day
#which means that the data is incorrect
#and -1 is not possible
```

```
final_dataset = final_dataset[(final_dataset['item_cnt_month'] > 0) &
                             ~(final_dataset['item_cnt_month'] < 307980)]
```

```
print(final_dataset.shape)
```

```
(2928487, 10)
```

```
[ ]: #outlier treatment

#checks for outliers in the item_cnt_month column
print("\n\nOutliers in item_cnt_month column:\n")
print(final_dataset[final_dataset['item_cnt_month'] > 1000])

#removes the outliers in the item_cnt_month column
final_dataset = final_dataset[final_dataset['item_cnt_month'] < 1000]

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Outliers in item_cnt_month column:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
2909818	28.10.2015	33	12	11373	0.908714	2169	
				item_name	category_id	\	
2909818	Delivery to the point of issue (Boxberry)				9		
				item_category_name		shop_name	
2909818	Delivery of goods			Internet-shop of emergency situations			

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00	1	
1	03.01.2013	0	25	2552	899.00	1	
3	06.01.2013	0	25	2554	1709.05	1	
4	15.01.2013	0	25	2555	1099.00	1	
5	10.01.2013	0	25	2564	349.00	1	
6	02.01.2013	0	25	2565	549.00	1	
7	04.01.2013	0	25	2572	239.00	1	
8	11.01.2013	0	25	2572	299.00	1	
9	03.01.2013	0	25	2573	299.00	3	
10	03.01.2013	0	25	2574	399.00	2	
11	05.01.2013	0	25	2574	399.00	1	
12	07.01.2013	0	25	2574	399.00	1	
13	08.01.2013	0	25	2574	399.00	2	
14	10.01.2013	0	25	2574	399.00	1	
15	11.01.2013	0	25	2574	399.00	2	
16	13.01.2013	0	25	2574	399.00	1	
17	16.01.2013	0	25	2574	399.00	1	
18	26.01.2013	0	25	2574	399.00	1	

19	27.01.2013	0	25	2574	399.00	1
20	09.01.2013	0	25	2593	279.00	1

				item_name	category_id	\
0				SCENE 2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light	LP		58	
3	DEEP PURPLE	Who Do You Think We Are	LP		58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).			56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...			59	
6		DEEP PURPLE Stormbringer (firms).			56	
7		DEFTONES Koi No Yakan			55	
8		DEFTONES Koi No Yakan			55	
9		DEL REY LANA Born To Die			55	
10	DEL REY LANA	Born To Die	The Paradise Editio...		55	
11	DEL REY LANA	Born To Die	The Paradise Editio...		55	
12	DEL REY LANA	Born To Die	The Paradise Editio...		55	
13	DEL REY LANA	Born To Die	The Paradise Editio...		55	
14	DEL REY LANA	Born To Die	The Paradise Editio...		55	
15	DEL REY LANA	Born To Die	The Paradise Editio...		55	
16	DEL REY LANA	Born To Die	The Paradise Editio...		55	
17	DEL REY LANA	Born To Die	The Paradise Editio...		55	
18	DEL REY LANA	Born To Die	The Paradise Editio...		55	
19	DEL REY LANA	Born To Die	The Paradise Editio...		55	
20		DEPECHE MODE	Music For The Masses		55	

	item_category_name	shop_name
0	Cinema - Blu-Ray	Yaroslavl TC" Altair "
1	Music - Vinyl	Moscow TEC" Atrium "
3	Music - Vinyl	Moscow TEC" Atrium "
4	Music - CD of corporate production	Moscow TEC" Atrium "
5	Music - Music video	Moscow TEC" Atrium "
6	Music - CD of corporate production	Moscow TEC" Atrium "
7	Music - CD of local production	Moscow TEC" Atrium "
8	Music - CD of local production	Moscow TEC" Atrium "
9	Music - CD of local production	Moscow TEC" Atrium "
10	Music - CD of local production	Moscow TEC" Atrium "
11	Music - CD of local production	Moscow TEC" Atrium "
12	Music - CD of local production	Moscow TEC" Atrium "
13	Music - CD of local production	Moscow TEC" Atrium "
14	Music - CD of local production	Moscow TEC" Atrium "
15	Music - CD of local production	Moscow TEC" Atrium "
16	Music - CD of local production	Moscow TEC" Atrium "
17	Music - CD of local production	Moscow TEC" Atrium "
18	Music - CD of local production	Moscow TEC" Atrium "
19	Music - CD of local production	Moscow TEC" Atrium "
20	Music - CD of local production	Moscow TEC" Atrium "

(2928485, 10)

```
[ ]: #deal with the incorrect data in the item_price column
#the item_price should not be negative
#the item_price should not be zero
#the item_price should not be greater than 100000

final_dataset = final_dataset[(final_dataset['item_price'] > 0) &
                             ~(final_dataset['item_price'] < 100000)]
```

```
[ ]: print(final_dataset.shape)
```

```
(2928483, 10)
```

```
[ ]: #handles special characters and formatting in the data set
final_dataset['item_name'] = final_dataset['item_name'].str.
    replace('[^A-Za-z0-9 - - ]+', ' ')
```

```
[ ]: print(final_dataset.shape)
```

```
(2928483, 10)
```

```
[ ]: #removes the noise in the item_name column
final_dataset['item_name'] = final_dataset['item_name'].str.replace(' ', '')
```

```
[ ]: print(final_dataset.head())
```

	date	month_num	shop_id	item_id	item_price	item_cnt_month	item_name	category_id	shop_name
0	02.01.2013	0	59	22154	999.00	1	SCENE 2012 (BD)	37	Yaroslavl TC" Altair "
1	03.01.2013	0	25	2552	899.00	1	DEEP PURPLE The House Of Blue Light LP	58	Moscow TEC" Atrium "
3	06.01.2013	0	25	2554	1709.05	1	DEEP PURPLE Who Do You Think We Are LP	58	Moscow TEC" Atrium "
4	15.01.2013	0	25	2555	1099.00	1	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	Moscow TEC" Atrium "
5	10.01.2013	0	25	2564	349.00	1	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	Moscow TEC" Atrium "

```
[ ]: #creates a new column called revenue
final_dataset['revenue'] = final_dataset['item_cnt_month'] * final_dataset['item_price']
```

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00		1
1	03.01.2013	0	25	2552	899.00		1
3	06.01.2013	0	25	2554	1709.05		1
4	15.01.2013	0	25	2555	1099.00		1
5	10.01.2013	0	25	2564	349.00		1
6	02.01.2013	0	25	2565	549.00		1
7	04.01.2013	0	25	2572	239.00		1
8	11.01.2013	0	25	2572	299.00		1
9	03.01.2013	0	25	2573	299.00		3
10	03.01.2013	0	25	2574	399.00		2
11	05.01.2013	0	25	2574	399.00		1
12	07.01.2013	0	25	2574	399.00		1
13	08.01.2013	0	25	2574	399.00		2
14	10.01.2013	0	25	2574	399.00		1
15	11.01.2013	0	25	2574	399.00		2
16	13.01.2013	0	25	2574	399.00		1
17	16.01.2013	0	25	2574	399.00		1
18	26.01.2013	0	25	2574	399.00		1
19	27.01.2013	0	25	2574	399.00		1
20	09.01.2013	0	25	2593	279.00		1

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE Stormbringer (firms).	56	
7	DEFTONES Koi No Yakan	55	
8	DEFTONES Koi No Yakan	55	
9	DEL REY LANA Born To Die	55	
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55	

```

18 DEL REY LANA Born To Die The Paradise Edition 2CD      55
19 DEL REY LANA Born To Die The Paradise Edition 2CD      55
20          DEPECHE MODE Music For The Masses            55

```

	item_category_name	shop_name	revenue	
0	Cinema - Blu-Ray	Yaroslavl TEC"	Altair "	999.00
1	Music - Vinyl	Moscow TEC"	Atrium "	899.00
3	Music - Vinyl	Moscow TEC"	Atrium "	1709.05
4	Music - CD of corporate production	Moscow TEC"	Atrium "	1099.00
5	Music - Music video	Moscow TEC"	Atrium "	349.00
6	Music - CD of corporate production	Moscow TEC"	Atrium "	549.00
7	Music - CD of local production	Moscow TEC"	Atrium "	239.00
8	Music - CD of local production	Moscow TEC"	Atrium "	299.00
9	Music - CD of local production	Moscow TEC"	Atrium "	897.00
10	Music - CD of local production	Moscow TEC"	Atrium "	798.00
11	Music - CD of local production	Moscow TEC"	Atrium "	399.00
12	Music - CD of local production	Moscow TEC"	Atrium "	399.00
13	Music - CD of local production	Moscow TEC"	Atrium "	798.00
14	Music - CD of local production	Moscow TEC"	Atrium "	399.00
15	Music - CD of local production	Moscow TEC"	Atrium "	798.00
16	Music - CD of local production	Moscow TEC"	Atrium "	399.00
17	Music - CD of local production	Moscow TEC"	Atrium "	399.00
18	Music - CD of local production	Moscow TEC"	Atrium "	399.00
19	Music - CD of local production	Moscow TEC"	Atrium "	399.00
20	Music - CD of local production	Moscow TEC"	Atrium "	279.00

(2928483, 11)

```

[ ]: #creates a new column called revenue_per_item
final_dataset['revenue_per_item'] = final_dataset['revenue'] / final_dataset['item_cnt_month']

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)

```

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00		1
1	03.01.2013	0	25	2552	899.00		1
3	06.01.2013	0	25	2554	1709.05		1
4	15.01.2013	0	25	2555	1099.00		1
5	10.01.2013	0	25	2564	349.00		1
6	02.01.2013	0	25	2565	549.00		1
7	04.01.2013	0	25	2572	239.00		1

8	11.01.2013	0	25	2572	299.00	1
9	03.01.2013	0	25	2573	299.00	3
10	03.01.2013	0	25	2574	399.00	2
11	05.01.2013	0	25	2574	399.00	1
12	07.01.2013	0	25	2574	399.00	1
13	08.01.2013	0	25	2574	399.00	2
14	10.01.2013	0	25	2574	399.00	1
15	11.01.2013	0	25	2574	399.00	2
16	13.01.2013	0	25	2574	399.00	1
17	16.01.2013	0	25	2574	399.00	1
18	26.01.2013	0	25	2574	399.00	1
19	27.01.2013	0	25	2574	399.00	1
20	09.01.2013	0	25	2593	279.00	1

		item_name	category_id	\
0		SCENE 2012 (BD)	37	
1		DEEP PURPLE The House Of Blue Light LP	58	
3		DEEP PURPLE Who Do You Think We Are LP	58	
4		DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5		DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6		DEEP PURPLE Stormbringer (firms).	56	
7		DEFTONES Koi No Yokan	55	
8		DEFTONES Koi No Yokan	55	
9		DEL REY LANA Born To Die	55	
10		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
18		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
19		DEL REY LANA Born To Die The Paradise Edition 2CD	55	
20		DEPECHE MODE Music For The Masses	55	

	item_category_name	shop_name	revenue	\
0	Cinema - Blu-Ray	Yaroslavl TC" Altair "	999.00	
1	Music - Vinyl	Moscow TEC" Atrium "	899.00	
3	Music - Vinyl	Moscow TEC" Atrium "	1709.05	
4	Music - CD of corporate production	Moscow TEC" Atrium "	1099.00	
5	Music - Music video	Moscow TEC" Atrium "	349.00	
6	Music - CD of corporate production	Moscow TEC" Atrium "	549.00	
7	Music - CD of local production	Moscow TEC" Atrium "	239.00	
8	Music - CD of local production	Moscow TEC" Atrium "	299.00	
9	Music - CD of local production	Moscow TEC" Atrium "	897.00	
10	Music - CD of local production	Moscow TEC" Atrium "	798.00	
11	Music - CD of local production	Moscow TEC" Atrium "	399.00	

```

12      Music - CD of local production    Moscow TEC" Atrium " 399.00
13      Music - CD of local production    Moscow TEC" Atrium " 798.00
14      Music - CD of local production    Moscow TEC" Atrium " 399.00
15      Music - CD of local production    Moscow TEC" Atrium " 798.00
16      Music - CD of local production    Moscow TEC" Atrium " 399.00
17      Music - CD of local production    Moscow TEC" Atrium " 399.00
18      Music - CD of local production    Moscow TEC" Atrium " 399.00
19      Music - CD of local production    Moscow TEC" Atrium " 399.00
20      Music - CD of local production    Moscow TEC" Atrium " 279.00

revenue_per_item
0          999.00
1          899.00
3         1709.05
4         1099.00
5          349.00
6          549.00
7          239.00
8          299.00
9          299.00
10         399.00
11         399.00
12         399.00
13         399.00
14         399.00
15         399.00
16         399.00
17         399.00
18         399.00
19         399.00
20         279.00
(2928483, 12)

```

```
[ ]: #checks whether the revenue_per_item column and revenue column are the same

if final_dataset['revenue_per_item'].equals(final_dataset['revenue']):
    # If they are the same, drop one of the columns
    final_dataset.drop(columns=['revenue_per_item'], inplace=True)

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

```
date  month_num  shop_id  item_id  item_price  item_cnt_month \
```

0	02.01.2013	0	59	22154	999.00	1
1	03.01.2013	0	25	2552	899.00	1
3	06.01.2013	0	25	2554	1709.05	1
4	15.01.2013	0	25	2555	1099.00	1
5	10.01.2013	0	25	2564	349.00	1
6	02.01.2013	0	25	2565	549.00	1
7	04.01.2013	0	25	2572	239.00	1
8	11.01.2013	0	25	2572	299.00	1
9	03.01.2013	0	25	2573	299.00	3
10	03.01.2013	0	25	2574	399.00	2
11	05.01.2013	0	25	2574	399.00	1
12	07.01.2013	0	25	2574	399.00	1
13	08.01.2013	0	25	2574	399.00	2
14	10.01.2013	0	25	2574	399.00	1
15	11.01.2013	0	25	2574	399.00	2
16	13.01.2013	0	25	2574	399.00	1
17	16.01.2013	0	25	2574	399.00	1
18	26.01.2013	0	25	2574	399.00	1
19	27.01.2013	0	25	2574	399.00	1
20	09.01.2013	0	25	2593	279.00	1

0		item_name	category_id	\
	SCENE	2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light LP	58	
3	DEEP PURPLE	Who Do You Think We Are LP	58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE	Stormbringer (firms).	56	
7	DEFTONES	Koi No Yakan	55	
8	DEFTONES	Koi No Yakan	55	
9	DEL REY LANA	Born To Die	55	
10	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
14	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
15	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
16	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
17	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
18	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
19	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
20	DEPECHE MODE	Music For The Masses	55	

0	item_category_name	shop_name	revenue	\
	Cinema - Blu-Ray	Yaroslavl TC"	Altair "	999.00
1	Music - Vinyl	Moscow TEC"	Atrium "	899.00
3	Music - Vinyl	Moscow TEC"	Atrium "	1709.05
4	Music - CD of corporate production	Moscow TEC"	Atrium "	1099.00

```

5           Music - Music video      Moscow TEC" Atrium "  349.00
6   Music - CD of corporate production Moscow TEC" Atrium "  549.00
7       Music - CD of local production Moscow TEC" Atrium "  239.00
8       Music - CD of local production Moscow TEC" Atrium "  299.00
9       Music - CD of local production Moscow TEC" Atrium "  897.00
10      Music - CD of local production Moscow TEC" Atrium "  798.00
11      Music - CD of local production Moscow TEC" Atrium "  399.00
12      Music - CD of local production Moscow TEC" Atrium "  399.00
13      Music - CD of local production Moscow TEC" Atrium "  798.00
14      Music - CD of local production Moscow TEC" Atrium "  399.00
15      Music - CD of local production Moscow TEC" Atrium "  798.00
16      Music - CD of local production Moscow TEC" Atrium "  399.00
17      Music - CD of local production Moscow TEC" Atrium "  399.00
18      Music - CD of local production Moscow TEC" Atrium "  399.00
19      Music - CD of local production Moscow TEC" Atrium "  399.00
20      Music - CD of local production Moscow TEC" Atrium "  279.00

revenue_per_item
0          999.00
1          899.00
3         1709.05
4         1099.00
5          349.00
6          549.00
7          239.00
8          299.00
9          299.00
10         399.00
11         399.00
12         399.00
13         399.00
14         399.00
15         399.00
16         399.00
17         399.00
18         399.00
19         399.00
20         279.00
(2928483, 12)

```

```
[ ]: #creates a new column called date num
final_dataset['date_num'] = final_dataset['date'].str[:2]

[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00	1	
1	03.01.2013	0	25	2552	899.00	1	
3	06.01.2013	0	25	2554	1709.05	1	
4	15.01.2013	0	25	2555	1099.00	1	
5	10.01.2013	0	25	2564	349.00	1	
6	02.01.2013	0	25	2565	549.00	1	
7	04.01.2013	0	25	2572	239.00	1	
8	11.01.2013	0	25	2572	299.00	1	
9	03.01.2013	0	25	2573	299.00	3	
10	03.01.2013	0	25	2574	399.00	2	
11	05.01.2013	0	25	2574	399.00	1	
12	07.01.2013	0	25	2574	399.00	1	
13	08.01.2013	0	25	2574	399.00	2	
14	10.01.2013	0	25	2574	399.00	1	
15	11.01.2013	0	25	2574	399.00	2	
16	13.01.2013	0	25	2574	399.00	1	
17	16.01.2013	0	25	2574	399.00	1	
18	26.01.2013	0	25	2574	399.00	1	
19	27.01.2013	0	25	2574	399.00	1	
20	09.01.2013	0	25	2593	279.00	1	

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE Stormbringer (firms).	56	
7	DEFTONES Koi No Yakan	55	
8	DEFTONES Koi No Yakan	55	
9	DEL REY LANA Born To Die	55	
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
20	DEPECHE MODE Music For The Masses	55	

item_category_name shop_name revenue \

0	Cinema - Blu-Ray	Yaroslavl TC"	Altair "	999.00
1	Music - Vinyl	Moscow TEC"	Atrium "	899.00
3	Music - Vinyl	Moscow TEC"	Atrium "	1709.05
4	Music - CD of corporate production	Moscow TEC"	Atrium "	1099.00
5	Music - Music video	Moscow TEC"	Atrium "	349.00
6	Music - CD of corporate production	Moscow TEC"	Atrium "	549.00
7	Music - CD of local production	Moscow TEC"	Atrium "	239.00
8	Music - CD of local production	Moscow TEC"	Atrium "	299.00
9	Music - CD of local production	Moscow TEC"	Atrium "	897.00
10	Music - CD of local production	Moscow TEC"	Atrium "	798.00
11	Music - CD of local production	Moscow TEC"	Atrium "	399.00
12	Music - CD of local production	Moscow TEC"	Atrium "	399.00
13	Music - CD of local production	Moscow TEC"	Atrium "	798.00
14	Music - CD of local production	Moscow TEC"	Atrium "	399.00
15	Music - CD of local production	Moscow TEC"	Atrium "	798.00
16	Music - CD of local production	Moscow TEC"	Atrium "	399.00
17	Music - CD of local production	Moscow TEC"	Atrium "	399.00
18	Music - CD of local production	Moscow TEC"	Atrium "	399.00
19	Music - CD of local production	Moscow TEC"	Atrium "	399.00
20	Music - CD of local production	Moscow TEC"	Atrium "	279.00

	revenue_per_item	date_num
0	999.00	02
1	899.00	03
3	1709.05	06
4	1099.00	15
5	349.00	10
6	549.00	02
7	239.00	04
8	299.00	11
9	299.00	03
10	399.00	03
11	399.00	05
12	399.00	07
13	399.00	08
14	399.00	10
15	399.00	11
16	399.00	13
17	399.00	16
18	399.00	26
19	399.00	27
20	279.00	09

(2928483, 13)

```
[ ]: #creates a new column called year num
final_dataset['year_num'] = final_dataset['date'].str[6:]
```

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	month_num	shop_id	item_id	item_price	item_cnt_month	\
0	02.01.2013	0	59	22154	999.00		1
1	03.01.2013	0	25	2552	899.00		1
3	06.01.2013	0	25	2554	1709.05		1
4	15.01.2013	0	25	2555	1099.00		1
5	10.01.2013	0	25	2564	349.00		1
6	02.01.2013	0	25	2565	549.00		1
7	04.01.2013	0	25	2572	239.00		1
8	11.01.2013	0	25	2572	299.00		1
9	03.01.2013	0	25	2573	299.00		3
10	03.01.2013	0	25	2574	399.00		2
11	05.01.2013	0	25	2574	399.00		1
12	07.01.2013	0	25	2574	399.00		1
13	08.01.2013	0	25	2574	399.00		2
14	10.01.2013	0	25	2574	399.00		1
15	11.01.2013	0	25	2574	399.00		2
16	13.01.2013	0	25	2574	399.00		1
17	16.01.2013	0	25	2574	399.00		1
18	26.01.2013	0	25	2574	399.00		1
19	27.01.2013	0	25	2574	399.00		1
20	09.01.2013	0	25	2593	279.00		1

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE Stormbringer (firms).	56	
7	DEFTONES Koi No Yakan	55	
8	DEFTONES Koi No Yakan	55	
9	DEL REY LANA Born To Die	55	
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55	

18	DEL REY LANA Born To Die The Paradise Edition 2CD	55
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55
20	DEPECHE MODE Music For The Masses	55

	item_category_name	shop_name	revenue	\
0	Cinema - Blu-Ray	Yaroslavl TEC"	Altair "	999.00
1	Music - Vinyl	Moscow TEC"	Atrium "	899.00
3	Music - Vinyl	Moscow TEC"	Atrium "	1709.05
4	Music - CD of corporate production	Moscow TEC"	Atrium "	1099.00
5	Music - Music video	Moscow TEC"	Atrium "	349.00
6	Music - CD of corporate production	Moscow TEC"	Atrium "	549.00
7	Music - CD of local production	Moscow TEC"	Atrium "	239.00
8	Music - CD of local production	Moscow TEC"	Atrium "	299.00
9	Music - CD of local production	Moscow TEC"	Atrium "	897.00
10	Music - CD of local production	Moscow TEC"	Atrium "	798.00
11	Music - CD of local production	Moscow TEC"	Atrium "	399.00
12	Music - CD of local production	Moscow TEC"	Atrium "	399.00
13	Music - CD of local production	Moscow TEC"	Atrium "	798.00
14	Music - CD of local production	Moscow TEC"	Atrium "	399.00
15	Music - CD of local production	Moscow TEC"	Atrium "	798.00
16	Music - CD of local production	Moscow TEC"	Atrium "	399.00
17	Music - CD of local production	Moscow TEC"	Atrium "	399.00
18	Music - CD of local production	Moscow TEC"	Atrium "	399.00
19	Music - CD of local production	Moscow TEC"	Atrium "	399.00
20	Music - CD of local production	Moscow TEC"	Atrium "	279.00

	revenue_per_item	date_num	year_num
0	999.00	02	2013
1	899.00	03	2013
3	1709.05	06	2013
4	1099.00	15	2013
5	349.00	10	2013
6	549.00	02	2013
7	239.00	04	2013
8	299.00	11	2013
9	299.00	03	2013
10	399.00	03	2013
11	399.00	05	2013
12	399.00	07	2013
13	399.00	08	2013
14	399.00	10	2013
15	399.00	11	2013
16	399.00	13	2013
17	399.00	16	2013
18	399.00	26	2013
19	399.00	27	2013
20	279.00	09	2013

(2928483, 14)

```
[ ]: print(final_dataset.shape)
print(final_dataset.info())

(2928483, 14)
<class 'pandas.core.frame.DataFrame'>
Index: 2928483 entries, 0 to 2935848
Data columns (total 14 columns):
 #   Column           Dtype  
--- 
 0   date             object  
 1   month_num        int64  
 2   shop_id          int64  
 3   item_id          int64  
 4   item_price       float64
 5   item_cnt_month   int64  
 6   item_name        object  
 7   category_id      int64  
 8   item_category_name object  
 9   shop_name         object  
 10  revenue          float64
 11  revenue_per_item float64
 12  date_num         object  
 13  year_num         object  
dtypes: float64(3), int64(5), object(6)
memory usage: 335.1+ MB
None
```

```
[ ]: # rearrange the columns
final_dataset = final_dataset[['date', 'date_num', 'year_num', 'month_num', ↴
    ↴'shop_id', 'shop_name', 'item_id', 'item_name', 'category_id', ↴
    ↴'item_category_name', 'item_price', 'item_cnt_month', 'revenue']]

print(final_dataset.shape)
print(final_dataset.info())
```

```
(2928483, 13)
<class 'pandas.core.frame.DataFrame'>
Index: 2928483 entries, 0 to 2935848
Data columns (total 13 columns):
 #   Column           Dtype  
--- 
 0   date             object  
 1   date_num         object  
 2   year_num         object  
 3   month_num        int64  
 4   shop_id          int64  
 5   shop_name         object  
 6   item_id          int64
```

```

7   item_name          object
8   category_id        int64
9   item_category_name object
10  item_price         float64
11  item_cnt_month    int64
12  revenue            float64
dtypes: float64(2), int64(5), object(6)
memory usage: 312.8+ MB
None

```

[]: #data profiling

```

#descriptive statistics
print("\n\nDescriptive statistics of final_dataset:\n")
print(final_dataset.describe())

```

Descriptive statistics of final_dataset:

	month_num	shop_id	item_id	category_id	item_price	\
count	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	
mean	1.456976e+01	3.300296e+01	1.020028e+04	4.001637e+01	8.893627e+02	
std	9.422952e+00	1.622543e+01	6.324391e+03	1.709809e+01	1.718155e+03	
min	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	7.000000e-02	
25%	7.000000e+00	2.200000e+01	4.477000e+03	2.800000e+01	2.490000e+02	
50%	1.400000e+01	3.100000e+01	9.355000e+03	4.000000e+01	3.990000e+02	
75%	2.300000e+01	4.700000e+01	1.569100e+04	5.500000e+01	9.990000e+02	
max	3.300000e+01	5.900000e+01	2.216900e+04	8.300000e+01	5.920000e+04	

	item_cnt_month	revenue
count	2.928483e+06	2.928483e+06
mean	1.247257e+00	1.164267e+03
std	2.217429e+00	5.684853e+03
min	1.000000e+00	7.000000e-02
25%	1.000000e+00	2.490000e+02
50%	1.000000e+00	4.490000e+02
75%	1.000000e+00	1.090000e+03
max	6.690000e+02	1.829990e+06

[]: #data enrichment

```
#creates a new column called month name
```

```

final_dataset['month_name'] = final_dataset['month_num'].replace({0: 'January', 1: 'February', 2: 'March', 3: 'April', 4: 'May', 5: 'June', 6: 'July', 7: 'August', 8: 'September', 9: 'October', 10: 'November', 11: 'December', 12: 'January', 13: 'February', 14: 'March', 15: 'April', 16: 'May', 17: 'June', 18: 'July', 19: 'August', 20: 'September', 21: 'October', 22: 'November', 23: 'December', 24: 'January', 25: 'February', 26: 'March', 27: 'April', 28: 'May', 29: 'June', 30: 'July', 31: 'August', 32: 'September', 33: 'October'})}

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)

```

Head of final_dataset:

	date	date_num	year_num	month_num	shop_id	shop_name	\
0	02.01.2013	02	2013	0	59	Yaroslavl TEC	" Altair "
1	03.01.2013	03	2013	0	25	Moscow TEC	" Atrium "
3	06.01.2013	06	2013	0	25	Moscow TEC	" Atrium "
4	15.01.2013	15	2013	0	25	Moscow TEC	" Atrium "
5	10.01.2013	10	2013	0	25	Moscow TEC	" Atrium "
6	02.01.2013	02	2013	0	25	Moscow TEC	" Atrium "
7	04.01.2013	04	2013	0	25	Moscow TEC	" Atrium "
8	11.01.2013	11	2013	0	25	Moscow TEC	" Atrium "
9	03.01.2013	03	2013	0	25	Moscow TEC	" Atrium "
10	03.01.2013	03	2013	0	25	Moscow TEC	" Atrium "
11	05.01.2013	05	2013	0	25	Moscow TEC	" Atrium "
12	07.01.2013	07	2013	0	25	Moscow TEC	" Atrium "
13	08.01.2013	08	2013	0	25	Moscow TEC	" Atrium "
14	10.01.2013	10	2013	0	25	Moscow TEC	" Atrium "
15	11.01.2013	11	2013	0	25	Moscow TEC	" Atrium "
16	13.01.2013	13	2013	0	25	Moscow TEC	" Atrium "
17	16.01.2013	16	2013	0	25	Moscow TEC	" Atrium "
18	26.01.2013	26	2013	0	25	Moscow TEC	" Atrium "
19	27.01.2013	27	2013	0	25	Moscow TEC	" Atrium "
20	09.01.2013	09	2013	0	25	Moscow TEC	" Atrium "

	item_id	item_name	category_id	\
0	22154	SCENE 2012 (BD)	37	
1	2552	DEEP PURPLE The House Of Blue Light LP	58	
3	2554	DEEP PURPLE Who Do You Think We Are LP	58	
4	2555	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	2564	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6	2565	DEEP PURPLE Stormbringer (firms).	56	
7	2572	DEFTONES Koi No Yokan	55	
8	2572	DEFTONES Koi No Yokan	55	
9	2573	DEL REY LANA Born To Die	55	

10	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
11	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
12	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
13	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
14	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
15	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
16	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
17	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
18	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
19	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
20	2593	DEPECHE MODE Music For The Masses	55

		item_category_name	item_price	item_cnt_month	revenue	\
0		Cinema - Blu-Ray	999.00	1	999.00	
1		Music - Vinyl	899.00	1	899.00	
3		Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production		1099.00	1	1099.00	
5		Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production		549.00	1	549.00	
7		Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production		299.00	1	299.00	
9		Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production		399.00	2	798.00	
11		Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production		399.00	1	399.00	
13		Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production		399.00	1	399.00	
15		Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production		399.00	1	399.00	
17		Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production		399.00	1	399.00	
19		Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production		279.00	1	279.00	

	month_name
0	January
1	January
3	January
4	January
5	January
6	January
7	January
8	January
9	January
10	January
11	January
12	January
13	January

```
14    January
15    January
16    January
17    January
18    January
19    January
20    January
(2928483, 14)
```

```
[ ]: #removes month_num column

final_dataset.drop(columns=['month_num'], inplace=True)

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

```
      date date_num year_num shop_id          shop_name item_id \
0  02.01.2013       02    2013      59  Yaroslavl TC" Altair "  22154
1  03.01.2013       03    2013      25    Moscow TEC" Atrium "  2552
3  06.01.2013       06    2013      25    Moscow TEC" Atrium "  2554
4  15.01.2013      15    2013      25    Moscow TEC" Atrium "  2555
5  10.01.2013      10    2013      25    Moscow TEC" Atrium "  2564
6  02.01.2013       02    2013      25    Moscow TEC" Atrium "  2565
7  04.01.2013       04    2013      25    Moscow TEC" Atrium "  2572
8  11.01.2013       11    2013      25    Moscow TEC" Atrium "  2572
9  03.01.2013       03    2013      25    Moscow TEC" Atrium "  2573
10 03.01.2013       03    2013      25    Moscow TEC" Atrium "  2574
11 05.01.2013       05    2013      25    Moscow TEC" Atrium "  2574
12 07.01.2013       07    2013      25    Moscow TEC" Atrium "  2574
13 08.01.2013       08    2013      25    Moscow TEC" Atrium "  2574
14 10.01.2013      10    2013      25    Moscow TEC" Atrium "  2574
15 11.01.2013      11    2013      25    Moscow TEC" Atrium "  2574
16 13.01.2013      13    2013      25    Moscow TEC" Atrium "  2574
17 16.01.2013      16    2013      25    Moscow TEC" Atrium "  2574
18 26.01.2013      26    2013      25    Moscow TEC" Atrium "  2574
19 27.01.2013      27    2013      25    Moscow TEC" Atrium "  2574
20 09.01.2013      09    2013      25    Moscow TEC" Atrium "  2593

                  item_name category_id \
0              SCENE 2012 (BD)           37
1  DEEP PURPLE The House Of Blue Light LP        58
3  DEEP PURPLE Who Do You Think We Are LP        58
4  DEEP PURPLE 30 Very Best Of 2CD (Businesses).  56
```

5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59
6	DEEP PURPLE Stormbringer (firms).	56
7	DEFTONES Koi No Yakan	55
8	DEFTONES Koi No Yakan	55
9	DEL REY LANA Born To Die	55
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55
20	DEPECHE MODE Music For The Masses	55
0	item_category_name item_price item_cnt_month revenue \	
1	Cinema - Blu-Ray 999.00 1 999.00	
3	Music - Vinyl 899.00 1 899.00	
4	Music - Vinyl 1709.05 1 1709.05	
5	Music - CD of corporate production 1099.00 1 1099.00	
6	Music - Music video 349.00 1 349.00	
7	Music - CD of corporate production 549.00 1 549.00	
8	Music - CD of local production 239.00 1 239.00	
9	Music - CD of local production 299.00 1 299.00	
10	Music - CD of local production 299.00 3 897.00	
11	Music - CD of local production 399.00 2 798.00	
12	Music - CD of local production 399.00 1 399.00	
13	Music - CD of local production 399.00 1 399.00	
14	Music - CD of local production 399.00 1 399.00	
15	Music - CD of local production 399.00 1 399.00	
16	Music - CD of local production 399.00 1 399.00	
17	Music - CD of local production 399.00 1 399.00	
18	Music - CD of local production 399.00 1 399.00	
19	Music - CD of local production 399.00 1 399.00	
20	Music - CD of local production 279.00 1 279.00	
0	month_name	
1	January	
3	January	
4	January	
5	January	
6	January	
7	January	
8	January	

```

9    January
10   January
11   January
12   January
13   January
14   January
15   January
16   January
17   January
18   January
19   January
20   January
(2928483, 13)

```

[]: #rearrange the columns

```

final_dataset = final_dataset[['date', 'date_num', 'month_name', 'year_num', \
                             'shop_id', 'shop_name', 'item_id', 'item_name', 'category_id', \
                             'item_category_name', 'item_price', 'item_cnt_month', 'revenue']]
print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)

```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	shop_name	\
0	02.01.2013	02	January	2013	59	Yaroslavl TC"	Altair "
1	03.01.2013	03	January	2013	25	Moscow TEC"	Atrium "
3	06.01.2013	06	January	2013	25	Moscow TEC"	Atrium "
4	15.01.2013	15	January	2013	25	Moscow TEC"	Atrium "
5	10.01.2013	10	January	2013	25	Moscow TEC"	Atrium "
6	02.01.2013	02	January	2013	25	Moscow TEC"	Atrium "
7	04.01.2013	04	January	2013	25	Moscow TEC"	Atrium "
8	11.01.2013	11	January	2013	25	Moscow TEC"	Atrium "
9	03.01.2013	03	January	2013	25	Moscow TEC"	Atrium "
10	03.01.2013	03	January	2013	25	Moscow TEC"	Atrium "
11	05.01.2013	05	January	2013	25	Moscow TEC"	Atrium "
12	07.01.2013	07	January	2013	25	Moscow TEC"	Atrium "
13	08.01.2013	08	January	2013	25	Moscow TEC"	Atrium "
14	10.01.2013	10	January	2013	25	Moscow TEC"	Atrium "
15	11.01.2013	11	January	2013	25	Moscow TEC"	Atrium "
16	13.01.2013	13	January	2013	25	Moscow TEC"	Atrium "
17	16.01.2013	16	January	2013	25	Moscow TEC"	Atrium "
18	26.01.2013	26	January	2013	25	Moscow TEC"	Atrium "
19	27.01.2013	27	January	2013	25	Moscow TEC"	Atrium "

20 09.01.2013 09 January 2013 25 Moscow TEC" Atrium "

	item_id	item_name	category_id	\
0	22154	SCENE 2012 (BD)	37	
1	2552	DEEP PURPLE The House Of Blue Light LP	58	
3	2554	DEEP PURPLE Who Do You Think We Are LP	58	
4	2555	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	2564	DEEP PURPLE Perihelion: Live In Concert DVD (C..	59	
6	2565	DEEP PURPLE Stormbringer (firms).	56	
7	2572	DEFTONES Koi No Yakan	55	
8	2572	DEFTONES Koi No Yakan	55	
9	2573	DEL REY LANA Born To Die	55	
10	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
18	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
19	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
20	2593	DEPECHE MODE Music For The Masses	55	

	item_category_name	item_price	item_cnt_month	revenue
0	Cinema - Blu-Ray	999.00	1	999.00
1	Music - Vinyl	899.00	1	899.00
3	Music - Vinyl	1709.05	1	1709.05
4	Music - CD of corporate production	1099.00	1	1099.00
5	Music - Music video	349.00	1	349.00
6	Music - CD of corporate production	549.00	1	549.00
7	Music - CD of local production	239.00	1	239.00
8	Music - CD of local production	299.00	1	299.00
9	Music - CD of local production	299.00	3	897.00
10	Music - CD of local production	399.00	2	798.00
11	Music - CD of local production	399.00	1	399.00
12	Music - CD of local production	399.00	1	399.00
13	Music - CD of local production	399.00	2	798.00
14	Music - CD of local production	399.00	1	399.00
15	Music - CD of local production	399.00	2	798.00
16	Music - CD of local production	399.00	1	399.00
17	Music - CD of local production	399.00	1	399.00
18	Music - CD of local production	399.00	1	399.00
19	Music - CD of local production	399.00	1	399.00
20	Music - CD of local production	279.00	1	279.00

(2928483, 13)

```
[ ]: #data binning

#found the bins using the following code
print(final_dataset['item_price'].max())
print(final_dataset['item_price'].min())

#create a new column called price range
final_dataset['price_range'] = pd.cut(final_dataset['item_price'], bins=[-1, ↵
    ↵100, 200, 300, 400, 500, 600, 700, 800, 900, 100000], labels=['0-100', ↵
    ↵'100-200', '200-300', '300-400', '400-500', '500-600', '600-700', '700-800', ↵
    ↵'800-900', '900-100000'])
```

59200.0

0.07

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	shop_name	\
0	02.01.2013	02	January	2013	59	Yaroslavl TEC	" Altair "
1	03.01.2013	03	January	2013	25	Moscow TEC	" Atrium "
3	06.01.2013	06	January	2013	25	Moscow TEC	" Atrium "
4	15.01.2013	15	January	2013	25	Moscow TEC	" Atrium "
5	10.01.2013	10	January	2013	25	Moscow TEC	" Atrium "
6	02.01.2013	02	January	2013	25	Moscow TEC	" Atrium "
7	04.01.2013	04	January	2013	25	Moscow TEC	" Atrium "
8	11.01.2013	11	January	2013	25	Moscow TEC	" Atrium "
9	03.01.2013	03	January	2013	25	Moscow TEC	" Atrium "
10	03.01.2013	03	January	2013	25	Moscow TEC	" Atrium "
11	05.01.2013	05	January	2013	25	Moscow TEC	" Atrium "
12	07.01.2013	07	January	2013	25	Moscow TEC	" Atrium "
13	08.01.2013	08	January	2013	25	Moscow TEC	" Atrium "
14	10.01.2013	10	January	2013	25	Moscow TEC	" Atrium "
15	11.01.2013	11	January	2013	25	Moscow TEC	" Atrium "
16	13.01.2013	13	January	2013	25	Moscow TEC	" Atrium "
17	16.01.2013	16	January	2013	25	Moscow TEC	" Atrium "
18	26.01.2013	26	January	2013	25	Moscow TEC	" Atrium "
19	27.01.2013	27	January	2013	25	Moscow TEC	" Atrium "
20	09.01.2013	09	January	2013	25	Moscow TEC	" Atrium "
	item_id				item_name	category_id	\
0	22154				SCENE 2012 (BD)	37	
1	2552				DEEP PURPLE The House Of Blue Light LP	58	

3	2554	DEEP PURPLE Who Do You Think We Are LP	58
4	2555	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56
5	2564	DEEP PURPLE Perihelion: Live In Concert DVD (C..	59
6	2565	DEEP PURPLE Stormbringer (firms).	56
7	2572	DEFTONES Koi No Yakan	55
8	2572	DEFTONES Koi No Yakan	55
9	2573	DEL REY LANA Born To Die	55
10	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
11	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
12	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
13	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
14	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
15	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
16	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
17	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
18	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
19	2574	DEL REY LANA Born To Die The Paradise Edition 2CD	55
20	2593	DEPECHE MODE Music For The Masses	55

		item_category_name	item_price	item_cnt_month	revenue	\
0		Cinema - Blu-Ray	999.00	1	999.00	
1		Music - Vinyl	899.00	1	899.00	
3		Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production		1099.00	1	1099.00	
5		Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production		549.00	1	549.00	
7		Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production		299.00	1	299.00	
9		Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production		399.00	2	798.00	
11		Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production		399.00	1	399.00	
13		Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production		399.00	1	399.00	
15		Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production		399.00	1	399.00	
17		Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production		399.00	1	399.00	
19		Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production		279.00	1	279.00	

	price_range
0	900-100000
1	800-900
3	900-100000
4	900-100000
5	300-400
6	500-600

```
7    200-300
8    200-300
9    200-300
10   300-400
11   300-400
12   300-400
13   300-400
14   300-400
15   300-400
16   300-400
17   300-400
18   300-400
19   300-400
20   200-300
(2928483, 14)
```

```
[ ]: #log transformation
```

```
#creates a new column called log_revenue
final_dataset['log_revenue'] = np.log(final_dataset['revenue'])
```

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	shop_name \
0	02.01.2013	02	January	2013	59	Yaroslavl TC" Altair "
1	03.01.2013	03	January	2013	25	Moscow TEC" Atrium "
3	06.01.2013	06	January	2013	25	Moscow TEC" Atrium "
4	15.01.2013	15	January	2013	25	Moscow TEC" Atrium "
5	10.01.2013	10	January	2013	25	Moscow TEC" Atrium "
6	02.01.2013	02	January	2013	25	Moscow TEC" Atrium "
7	04.01.2013	04	January	2013	25	Moscow TEC" Atrium "
8	11.01.2013	11	January	2013	25	Moscow TEC" Atrium "
9	03.01.2013	03	January	2013	25	Moscow TEC" Atrium "
10	03.01.2013	03	January	2013	25	Moscow TEC" Atrium "
11	05.01.2013	05	January	2013	25	Moscow TEC" Atrium "
12	07.01.2013	07	January	2013	25	Moscow TEC" Atrium "
13	08.01.2013	08	January	2013	25	Moscow TEC" Atrium "
14	10.01.2013	10	January	2013	25	Moscow TEC" Atrium "
15	11.01.2013	11	January	2013	25	Moscow TEC" Atrium "
16	13.01.2013	13	January	2013	25	Moscow TEC" Atrium "
17	16.01.2013	16	January	2013	25	Moscow TEC" Atrium "
18	26.01.2013	26	January	2013	25	Moscow TEC" Atrium "

19	27.01.2013	27	January	2013	25	Moscow TEC"	Atrium "
20	09.01.2013	09	January	2013	25	Moscow TEC"	Atrium "

	item_id		item_name	category_id	\
0	22154		SCENE 2012 (BD)	37	
1	2552	DEEP PURPLE	The House Of Blue Light LP	58	
3	2554	DEEP PURPLE	Who Do You Think We Are LP	58	
4	2555	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56	
5	2564	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59	
6	2565		DEEP PURPLE Stormbringer (firms).	56	
7	2572		DEFTONES Koi No Yakan	55	
8	2572		DEFTONES Koi No Yakan	55	
9	2573		DEL REY LANA Born To Die	55	
10	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
11	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
12	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
13	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
14	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
15	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
16	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
17	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
18	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
19	2574	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
20	2593		DEPECHE MODE Music For The Masses	55	

	item_category_name	item_price	item_cnt_month	revenue	\
0	Cinema - Blu-Ray	999.00	1	999.00	
1	Music - Vinyl	899.00	1	899.00	
3	Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production	1099.00	1	1099.00	
5	Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production	549.00	1	549.00	
7	Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production	299.00	1	299.00	
9	Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production	399.00	2	798.00	
11	Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production	399.00	1	399.00	
13	Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production	399.00	1	399.00	
15	Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production	399.00	1	399.00	
17	Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production	399.00	1	399.00	
19	Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production	279.00	1	279.00	

price_range log_revenue

```

0    900-100000    6.906755
1      800-900    6.801283
3    900-100000    7.443693
4    900-100000    7.002156
5      300-400    5.855072
6      500-600    6.308098
7      200-300    5.476464
8      200-300    5.700444
9      200-300    6.799056
10     300-400    6.682109
11     300-400    5.988961
12     300-400    5.988961
13     300-400    6.682109
14     300-400    5.988961
15     300-400    6.682109
16     300-400    5.988961
17     300-400    5.988961
18     300-400    5.988961
19     300-400    5.988961
20     200-300    5.631212
(2928483, 15)

```

```
[ ]: #encoding

#encodes the year_num column to 0, 1, 2

final_dataset['year_num'] = final_dataset['year_num'].replace({'2013': 0,
                                                               '2014': 1, '2015': 2})

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	\
0	02.01.2013	02	January	0	59	
1	03.01.2013	03	January	0	25	
3	06.01.2013	06	January	0	25	
4	15.01.2013	15	January	0	25	
5	10.01.2013	10	January	0	25	
6	02.01.2013	02	January	0	25	
7	04.01.2013	04	January	0	25	
8	11.01.2013	11	January	0	25	
9	03.01.2013	03	January	0	25	

10	03.01.2013	03	January	0	25
11	05.01.2013	05	January	0	25
12	07.01.2013	07	January	0	25
13	08.01.2013	08	January	0	25
14	10.01.2013	10	January	0	25
15	11.01.2013	11	January	0	25
16	13.01.2013	13	January	0	25
17	16.01.2013	16	January	0	25
18	26.01.2013	26	January	0	25
19	27.01.2013	27	January	0	25
20	09.01.2013	09	January	0	25

		shop_name	item_id	\
0	Yaroslavl TEC"	Altair "	22154	
1	Moscow TEC"	Atrium "	2552	
3	Moscow TEC"	Atrium "	2554	
4	Moscow TEC"	Atrium "	2555	
5	Moscow TEC"	Atrium "	2564	
6	Moscow TEC"	Atrium "	2565	
7	Moscow TEC"	Atrium "	2572	
8	Moscow TEC"	Atrium "	2572	
9	Moscow TEC"	Atrium "	2573	
10	Moscow TEC"	Atrium "	2574	
11	Moscow TEC"	Atrium "	2574	
12	Moscow TEC"	Atrium "	2574	
13	Moscow TEC"	Atrium "	2574	
14	Moscow TEC"	Atrium "	2574	
15	Moscow TEC"	Atrium "	2574	
16	Moscow TEC"	Atrium "	2574	
17	Moscow TEC"	Atrium "	2574	
18	Moscow TEC"	Atrium "	2574	
19	Moscow TEC"	Atrium "	2574	
20	Moscow TEC"	Atrium "	2593	

		item_name	category_id	\
0		SCENE 2012 (BD)	37	
1	DEEP PURPLE	The House Of Blue Light LP	58	
3	DEEP PURPLE	Who Do You Think We Are LP	58	
4	DEEP PURPLE	30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE	Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE	Stormbringer (firms).	56	
7	DEFTONES	Koi No Yakan	55	
8	DEFTONES	Koi No Yakan	55	
9	DEL REY LANA	Born To Die	55	
10	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA	Born To Die The Paradise Edition 2CD	55	

14	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55			
20	DEPECHE MODE Music For The Masses	55			
0	item_category_name	item_price	item_cnt_month	revenue	\
0	Cinema - Blu-Ray	999.00	1	999.00	
1	Music - Vinyl	899.00	1	899.00	
3	Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production	1099.00	1	1099.00	
5	Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production	549.00	1	549.00	
7	Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production	299.00	1	299.00	
9	Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production	399.00	2	798.00	
11	Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production	399.00	1	399.00	
13	Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production	399.00	1	399.00	
15	Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production	399.00	1	399.00	
17	Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production	399.00	1	399.00	
19	Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production	279.00	1	279.00	
0	price_range	log_revenue			
0	900-100000	6.906755			
1	800-900	6.801283			
3	900-100000	7.443693			
4	900-100000	7.002156			
5	300-400	5.855072			
6	500-600	6.308098			
7	200-300	5.476464			
8	200-300	5.700444			
9	200-300	6.799056			
10	300-400	6.682109			
11	300-400	5.988961			
12	300-400	5.988961			
13	300-400	6.682109			
14	300-400	5.988961			
15	300-400	6.682109			
16	300-400	5.988961			
17	300-400	5.988961			

```
18      300-400    5.988961
19      300-400    5.988961
20      200-300    5.631212
(2928483, 15)
```

```
[ ]: #grouping and aggregation
```

```
#grouping the data set by shop_id and year_num and aggregating the
↪item_cnt_month column using sum

grouped_by_shop_id_and_year_num = final_dataset.groupby(['shop_id', ↪
↪'year_num']).agg({'item_cnt_month': 'sum'})

print("\n\nHead of grouped_by_shop_id_and_year_num:\n")
print(grouped_by_shop_id_and_year_num.head(60))
print(grouped_by_shop_id_and_year_num.shape)
```

Head of grouped_by_shop_id_and_year_num:

		item_cnt_month
shop_id	year_num	
0	0	11705
1	0	6311
2	0	9989
	1	12247
	2	8470
3	0	10242
	1	11039
	2	7194
4	0	19054
	1	15909
	2	9106
5	0	14717
	1	17041
	2	11089
6	0	46707
	1	35496
	2	18554
7	0	28218
	1	24530
	2	14411
8	0	3602
9	0	6531
	1	6155
	2	3188
10	0	11132

	1	8859
	2	4532
11	2	572
12	0	19494
	1	26379
	2	24947
13	0	13529
	1	6234
14	0	20114
	1	16003
	2	10397
15	0	32339
	1	24854
	2	14138
16	0	30250
	1	19569
	2	11953
17	0	9913
	1	15308
	2	697
18	0	33632
	1	18815
	2	13174
19	0	32860
	1	25353
	2	15455
20	1	3262
	2	2611
21	0	25258
	1	25167
	2	18259
22	0	26943
	1	20341
	2	13050
23	0	7722
	(159, 1)	

```
[ ]: #creates a new column called scaled_revenue
final_dataset['scaled_revenue'] = (final_dataset['revenue'] - final_dataset['revenue'].min()) / (final_dataset['revenue'].max() - final_dataset['revenue'].min())

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	\
0	02.01.2013	02	January	0	59	
1	03.01.2013	03	January	0	25	
3	06.01.2013	06	January	0	25	
4	15.01.2013	15	January	0	25	
5	10.01.2013	10	January	0	25	
6	02.01.2013	02	January	0	25	
7	04.01.2013	04	January	0	25	
8	11.01.2013	11	January	0	25	
9	03.01.2013	03	January	0	25	
10	03.01.2013	03	January	0	25	
11	05.01.2013	05	January	0	25	
12	07.01.2013	07	January	0	25	
13	08.01.2013	08	January	0	25	
14	10.01.2013	10	January	0	25	
15	11.01.2013	11	January	0	25	
16	13.01.2013	13	January	0	25	
17	16.01.2013	16	January	0	25	
18	26.01.2013	26	January	0	25	
19	27.01.2013	27	January	0	25	
20	09.01.2013	09	January	0	25	

	shop_name	item_id	\
0	Yaroslavl TC" Altair "	22154	
1	Moscow TEC" Atrium "	2552	
3	Moscow TEC" Atrium "	2554	
4	Moscow TEC" Atrium "	2555	
5	Moscow TEC" Atrium "	2564	
6	Moscow TEC" Atrium "	2565	
7	Moscow TEC" Atrium "	2572	
8	Moscow TEC" Atrium "	2572	
9	Moscow TEC" Atrium "	2573	
10	Moscow TEC" Atrium "	2574	
11	Moscow TEC" Atrium "	2574	
12	Moscow TEC" Atrium "	2574	
13	Moscow TEC" Atrium "	2574	
14	Moscow TEC" Atrium "	2574	
15	Moscow TEC" Atrium "	2574	
16	Moscow TEC" Atrium "	2574	
17	Moscow TEC" Atrium "	2574	
18	Moscow TEC" Atrium "	2574	
19	Moscow TEC" Atrium "	2574	
20	Moscow TEC" Atrium "	2593	

	item_name	category_id	\
0	SCENE 2012 (BD)	37	

1	DEEP PURPLE The House Of Blue Light LP	58
3	DEEP PURPLE Who Do You Think We Are LP	58
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56
5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59
6	DEEP PURPLE Stormbringer (firms).	56
7	DEFTONES Koi No Yakan	55
8	DEFTONES Koi No Yakan	55
9	DEL REY LANA Born To Die	55
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55
20	DEPECHE MODE Music For The Masses	55

	item_category_name	item_price	item_cnt_month	revenue	\
0	Cinema - Blu-Ray	999.00	1	999.00	
1	Music - Vinyl	899.00	1	899.00	
3	Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production	1099.00	1	1099.00	
5	Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production	549.00	1	549.00	
7	Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production	299.00	1	299.00	
9	Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production	399.00	2	798.00	
11	Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production	399.00	1	399.00	
13	Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production	399.00	1	399.00	
15	Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production	399.00	1	399.00	
17	Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production	399.00	1	399.00	
19	Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production	279.00	1	279.00	

	price_range	log_revenue	scaled_revenue
0	900-100000	6.906755	0.000546
1	800-900	6.801283	0.000491
3	900-100000	7.443693	0.000934
4	900-100000	7.002156	0.000601
5	300-400	5.855072	0.000191

```

6      500-600    6.308098    0.000300
7      200-300    5.476464    0.000131
8      200-300    5.700444    0.000163
9      200-300    6.799056    0.000490
10     300-400    6.682109    0.000436
11     300-400    5.988961    0.000218
12     300-400    5.988961    0.000218
13     300-400    6.682109    0.000436
14     300-400    5.988961    0.000218
15     300-400    6.682109    0.000436
16     300-400    5.988961    0.000218
17     300-400    5.988961    0.000218
18     300-400    5.988961    0.000218
19     300-400    5.988961    0.000218
20     200-300    5.631212    0.000152
(2928483, 16)

```

[]: #change month_name column to numeric

```

final_dataset['month_name'] = final_dataset['month_name'].replace({'January':1,
                                                               'February': 2, 'March': 3, 'April': 4, 'May': 5, 'June':6, 'July': 7,
                                                               'August': 8, 'September': 9, 'October': 10, 'November':11, 'December': 12})

print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)

```

Head of final_dataset:

	date	date_num	month_name	year_num	shop_id	\
0	02.01.2013	02	1	0	59	
1	03.01.2013	03	1	0	25	
3	06.01.2013	06	1	0	25	
4	15.01.2013	15	1	0	25	
5	10.01.2013	10	1	0	25	
6	02.01.2013	02	1	0	25	
7	04.01.2013	04	1	0	25	
8	11.01.2013	11	1	0	25	
9	03.01.2013	03	1	0	25	
10	03.01.2013	03	1	0	25	
11	05.01.2013	05	1	0	25	
12	07.01.2013	07	1	0	25	
13	08.01.2013	08	1	0	25	
14	10.01.2013	10	1	0	25	
15	11.01.2013	11	1	0	25	
16	13.01.2013	13	1	0	25	

17	16.01.2013	16	1	0	25
18	26.01.2013	26	1	0	25
19	27.01.2013	27	1	0	25
20	09.01.2013	09	1	0	25

	shop_name	item_id	\
0	Yaroslavl TC" Altair "	22154	
1	Moscow TEC" Atrium "	2552	
3	Moscow TEC" Atrium "	2554	
4	Moscow TEC" Atrium "	2555	
5	Moscow TEC" Atrium "	2564	
6	Moscow TEC" Atrium "	2565	
7	Moscow TEC" Atrium "	2572	
8	Moscow TEC" Atrium "	2572	
9	Moscow TEC" Atrium "	2573	
10	Moscow TEC" Atrium "	2574	
11	Moscow TEC" Atrium "	2574	
12	Moscow TEC" Atrium "	2574	
13	Moscow TEC" Atrium "	2574	
14	Moscow TEC" Atrium "	2574	
15	Moscow TEC" Atrium "	2574	
16	Moscow TEC" Atrium "	2574	
17	Moscow TEC" Atrium "	2574	
18	Moscow TEC" Atrium "	2574	
19	Moscow TEC" Atrium "	2574	
20	Moscow TEC" Atrium "	2593	

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	
5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59	
6	DEEP PURPLE Stormbringer (firms).	56	
7	DEFTONES Koi No Yakan	55	
8	DEFTONES Koi No Yakan	55	
9	DEL REY LANA Born To Die	55	
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55	
20	DEPECHE MODE Music For The Masses	55	

	item_category_name	item_price	item_cnt_month	revenue	\
0	Cinema - Blu-Ray	999.00	1	999.00	
1	Music - Vinyl	899.00	1	899.00	
3	Music - Vinyl	1709.05	1	1709.05	
4	Music - CD of corporate production	1099.00	1	1099.00	
5	Music - Music video	349.00	1	349.00	
6	Music - CD of corporate production	549.00	1	549.00	
7	Music - CD of local production	239.00	1	239.00	
8	Music - CD of local production	299.00	1	299.00	
9	Music - CD of local production	299.00	3	897.00	
10	Music - CD of local production	399.00	2	798.00	
11	Music - CD of local production	399.00	1	399.00	
12	Music - CD of local production	399.00	1	399.00	
13	Music - CD of local production	399.00	2	798.00	
14	Music - CD of local production	399.00	1	399.00	
15	Music - CD of local production	399.00	2	798.00	
16	Music - CD of local production	399.00	1	399.00	
17	Music - CD of local production	399.00	1	399.00	
18	Music - CD of local production	399.00	1	399.00	
19	Music - CD of local production	399.00	1	399.00	
20	Music - CD of local production	279.00	1	279.00	
	price_range	log_revenue	scaled_revenue		
0	900-100000	6.906755	0.000546		
1	800-900	6.801283	0.000491		
3	900-100000	7.443693	0.000934		
4	900-100000	7.002156	0.000601		
5	300-400	5.855072	0.000191		
6	500-600	6.308098	0.000300		
7	200-300	5.476464	0.000131		
8	200-300	5.700444	0.000163		
9	200-300	6.799056	0.000490		
10	300-400	6.682109	0.000436		
11	300-400	5.988961	0.000218		
12	300-400	5.988961	0.000218		
13	300-400	6.682109	0.000436		
14	300-400	5.988961	0.000218		
15	300-400	6.682109	0.000436		
16	300-400	5.988961	0.000218		
17	300-400	5.988961	0.000218		
18	300-400	5.988961	0.000218		
19	300-400	5.988961	0.000218		
20	200-300	5.631212	0.000152		

(2928483, 16)

1.4 Data Exploration & Analysis

```
[ ]: #correlation
```

```
numeric_columns = final_dataset.select_dtypes(include=['number'])
print("\n\nCorrelation of final_dataset:\n")
print(numeric_columns.corr())
```

Correlation of final_dataset:

```
month_name      month_name    year_num     shop_id     item_id     category_id \
month_name      1.000000   -0.134131   0.020646   -0.002103   -0.001055
year_num        -0.134131   1.000000   0.011835   0.010486   0.030406
shop_id         0.020646   0.011835   1.000000   0.029344   0.019339
item_id         -0.002103   0.010486   0.029344   1.000000   0.369384
category_id    -0.001055   0.030406   0.019339   0.369384   1.000000
item_price      0.044361   0.080574   -0.023966  -0.134759  -0.255033
item_cnt_month  0.021007   0.002510   -0.005660  0.018896   0.016922
revenue         0.030326   0.029541   -0.013540  -0.063423  -0.117860
log_revenue     0.057248   0.096145   -0.039415  -0.310144  -0.342896
scaled_revenue  0.030326   0.029541   -0.013540  -0.063423  -0.117860

item_price      item_price    item_cnt_month  revenue     log_revenue \
month_name      0.044361   0.021007   0.030326   0.057248
year_num        0.080574   0.002510   0.029541   0.096145
shop_id         -0.023966  -0.005660  -0.013540  -0.039415
item_id         -0.134759  0.018896   -0.063423  -0.310144
category_id    -0.255033  0.016922   -0.117860  -0.342896
item_price      1.000000   0.014437   0.436314   0.601371
item_cnt_month  0.014437   1.000000   0.481568   0.115253
revenue         0.436314   0.481568   1.000000   0.331392
log_revenue     0.601371   0.115253   0.331392   1.000000
scaled_revenue  0.436314   0.481568   1.000000   0.331392

scaled_revenue
```

month_name	year_num	shop_id	item_id	category_id	item_price	item_cnt_month	revenue	log_revenue	scaled_revenue
month_name	1.000000	-0.134131	0.020646	-0.002103	-0.001055	0.021007	0.030326	0.057248	0.030326
year_num	-0.134131	1.000000	0.011835	0.010486	0.030406	0.002510	0.029541	0.096145	0.029541
shop_id	0.020646	0.011835	1.000000	0.029344	0.019339	-0.005660	-0.013540	-0.039415	-0.013540
item_id	-0.002103	0.010486	0.029344	1.000000	0.369384	0.018896	-0.063423	-0.310144	-0.063423
category_id	-0.001055	0.030406	0.019339	0.369384	1.000000	0.016922	-0.117860	-0.342896	-0.117860
item_price	0.044361	0.080574	-0.023966	-0.134759	-0.255033	1.000000	0.436314	0.601371	0.436314
item_cnt_month	0.021007	0.002510	-0.005660	-0.013540	-0.039415	0.014437	1.000000	0.115253	0.481568
revenue	0.030326	0.029541	-0.013540	-0.063423	-0.117860	0.014437	0.481568	1.000000	0.481568
log_revenue	0.057248	0.096145	-0.039415	-0.310144	-0.342896	0.016922	0.331392	1.000000	0.331392
scaled_revenue	0.030326	0.029541	-0.013540	-0.063423	-0.117860	0.016922	0.331392	1.000000	1.000000

```
[ ]: #checks for missing values
print("\n\nMissing values in final_dataset:\n")
print(final_dataset.isnull().sum())

#checks for null values
print("\n\nNull values in final_dataset:\n")
print(final_dataset.isnull().sum())
```

Missing values in final_dataset:

date	0
date_num	0
month_name	0
year_num	0
shop_id	0
shop_name	0
item_id	0
item_name	0
category_id	0
item_category_name	0
item_price	0
item_cnt_month	0
revenue	0
price_range	0
log_revenue	0
scaled_revenue	0
dtype: int64	

Null values in final_dataset:

date	0
date_num	0
month_name	0
year_num	0
shop_id	0
shop_name	0
item_id	0
item_name	0
category_id	0
item_category_name	0
item_price	0
item_cnt_month	0
revenue	0
price_range	0
log_revenue	0

```
scaled_revenue      0  
dtype: int64
```

```
[ ]: #Descriptive analytics
```

```
# Summary Statistics  
print("\nDescriptive statistics of final_dataset:")  
print(final_dataset.describe())
```

```
Descriptive statistics of final_dataset:
```

	month_name	year_num	shop_id	item_id	category_id	\
count	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	
mean	6.248408e+00	7.767790e-01	3.300296e+01	1.020028e+04	4.001637e+01	
std	3.535921e+00	7.684598e-01	1.622543e+01	6.324391e+03	1.709809e+01	
min	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	
25%	3.000000e+00	0.000000e+00	2.200000e+01	4.477000e+03	2.800000e+01	
50%	6.000000e+00	1.000000e+00	3.100000e+01	9.355000e+03	4.000000e+01	
75%	9.000000e+00	1.000000e+00	4.700000e+01	1.569100e+04	5.500000e+01	
max	1.200000e+01	2.000000e+00	5.900000e+01	2.216900e+04	8.300000e+01	

	item_price	item_cnt_month	revenue	log_revenue	\
count	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	
mean	8.893627e+02	1.247257e+00	1.164267e+03	6.254676e+00	
std	1.718155e+03	2.217429e+00	5.684853e+03	1.171779e+00	
min	7.000000e-02	1.000000e+00	7.000000e-02	-2.659260e+00	
25%	2.490000e+02	1.000000e+00	2.490000e+02	5.517453e+00	
50%	3.990000e+02	1.000000e+00	4.490000e+02	6.107023e+00	
75%	9.990000e+02	1.000000e+00	1.090000e+03	6.993933e+00	
max	5.920000e+04	6.690000e+02	1.829990e+06	1.441982e+01	

	scaled_revenue
count	2.928483e+06
mean	6.361770e-04
std	3.106494e-03
min	0.000000e+00
25%	1.360281e-04
50%	2.453183e-04
75%	5.955934e-04
max	1.000000e+00

```
[ ]: #seasonality analysis
```

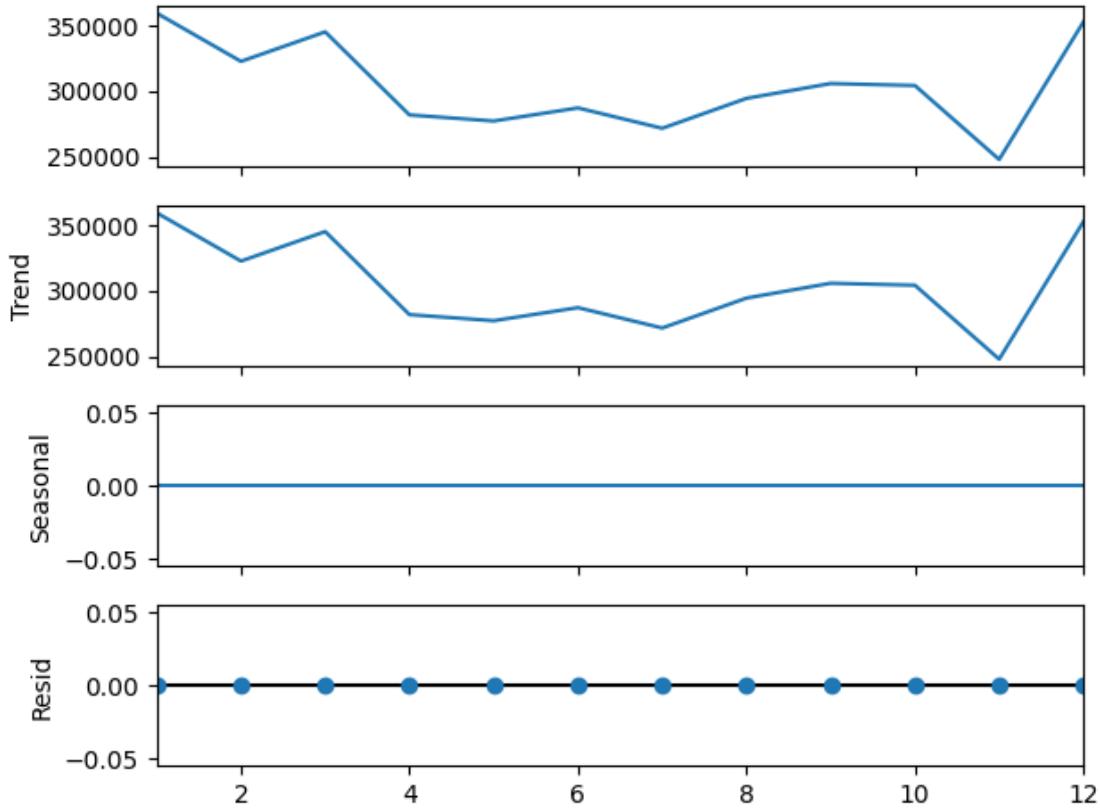
```
grouped_by_month_name = final_dataset.groupby(['month_name']).  
    agg({'item_cnt_month': 'sum'})  
  
print("\n\nHead of grouped_by_month_name:\n")  
print(grouped_by_month_name)
```

```
print(grouped_by_month_name.shape)
```

Head of grouped_by_month_name:

```
      item_cnt_month  
month_name  
1                  359391  
2                  322775  
3                  345171  
4                  282131  
5                  277519  
6                  287443  
7                  271985  
8                  294692  
9                  305967  
10                 304453  
11                 248320  
12                 352723  
(12, 1)
```

```
[ ]: #performing seasonal decomposition  
decomposition = sm.tsa.seasonal_decompose(grouped_by_month_name, model='additive', period=1)  
  
#plotting the seasonal decomposition  
fig = decomposition.plot()  
plt.show()  
  
#plotting the item_cnt_month column  
plt.figure(figsize=(20, 10))  
plt.plot(final_dataset['item_cnt_month'])  
plt.title('Item Count Per Month')  
plt.xlabel('Month')  
plt.ylabel('Item Count')  
plt.show()
```



```
[ ]: #regulatory analytics

grouped_by_shop_id_and_year_num = final_dataset.groupby(['shop_id', 'year_num']).agg({'item_cnt_month': 'sum'})

print("\n\nHead of grouped_by_shop_id_and_year_num:\n")
print(grouped_by_shop_id_and_year_num.head(60))
```

Head of grouped_by_shop_id_and_year_num:

shop_id	year_num	item_cnt_month
0	0	11705
1	0	6311
2	0	9989
	1	12247
	2	8470
3	0	10242
	1	11039
	2	7194
4	0	19054
	1	15909
	2	9106
5	0	14717
	1	17041
	2	11089
6	0	46707
	1	35496
	2	18554
7	0	28218
	1	24530
	2	14411
8	0	3602
9	0	6531
	1	6155
	2	3188
10	0	11132
	1	8859
	2	4532
11	2	572
12	0	19494
	1	26379
	2	24947
13	0	13529
	1	6234
14	0	20114

```

1          16003
2          10397
15         0        32339
           1        24854
           2        14138
16         0        30250
           1        19569
           2        11953
17         0        9913
           1        15308
           2        697
18         0        33632
           1        18815
           2        13174
19         0        32860
           1        25353
           2        15455
20         1        3262
           2        2611
21         0        25258
           1        25167
           2        18259
22         0        26943
           1        20341
           2        13050
23         0        7722

```

[]: #Variable Identification

```

# Identify numerical and categorical variables
numerical_vars = final_dataset.select_dtypes(include=['int64', 'float64']).
    ↪columns
categorical_vars = final_dataset.select_dtypes(include=['object', 'category']).
    ↪columns

# Print the list of numerical and categorical variables
print("Numerical Variables:")
print(numerical_vars)

print("\nCategorical Variables:")
print(categorical_vars)

```

Numerical Variables:

```

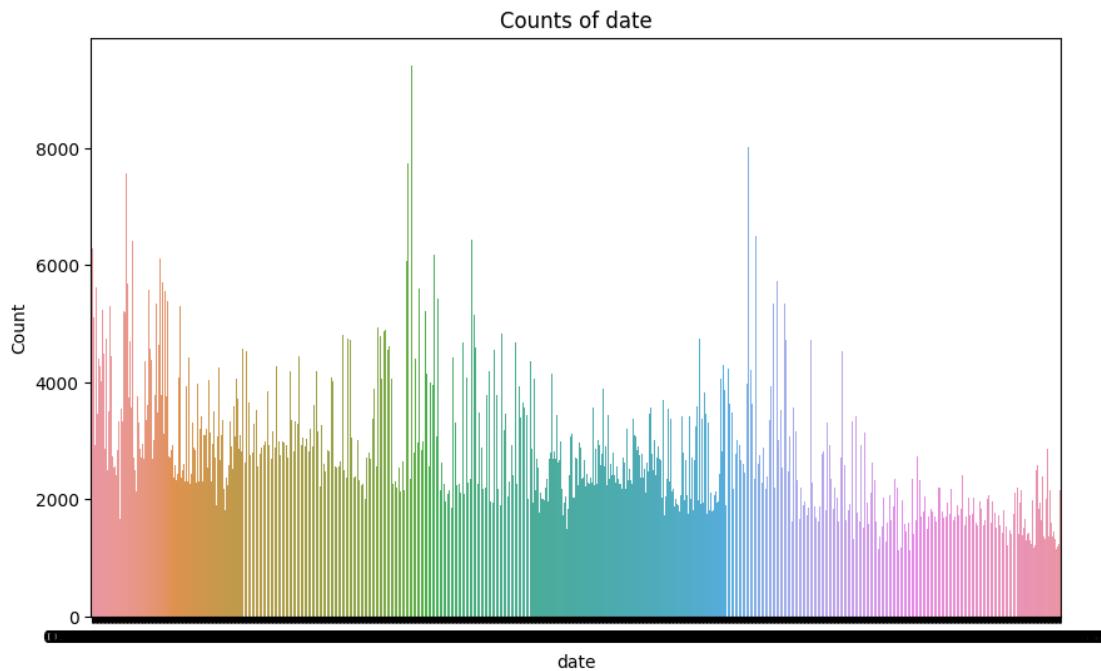
Index(['month_name', 'year_num', 'shop_id', 'item_id', 'category_id',
       'item_price', 'item_cnt_month', 'revenue', 'log_revenue',
       'scaled_revenue'],
      dtype='object')

```

```
Categorical Variables:  
Index(['date', 'date_num', 'shop_name', 'item_name', 'item_category_name',  
       'price_range'],  
      dtype='object')
```

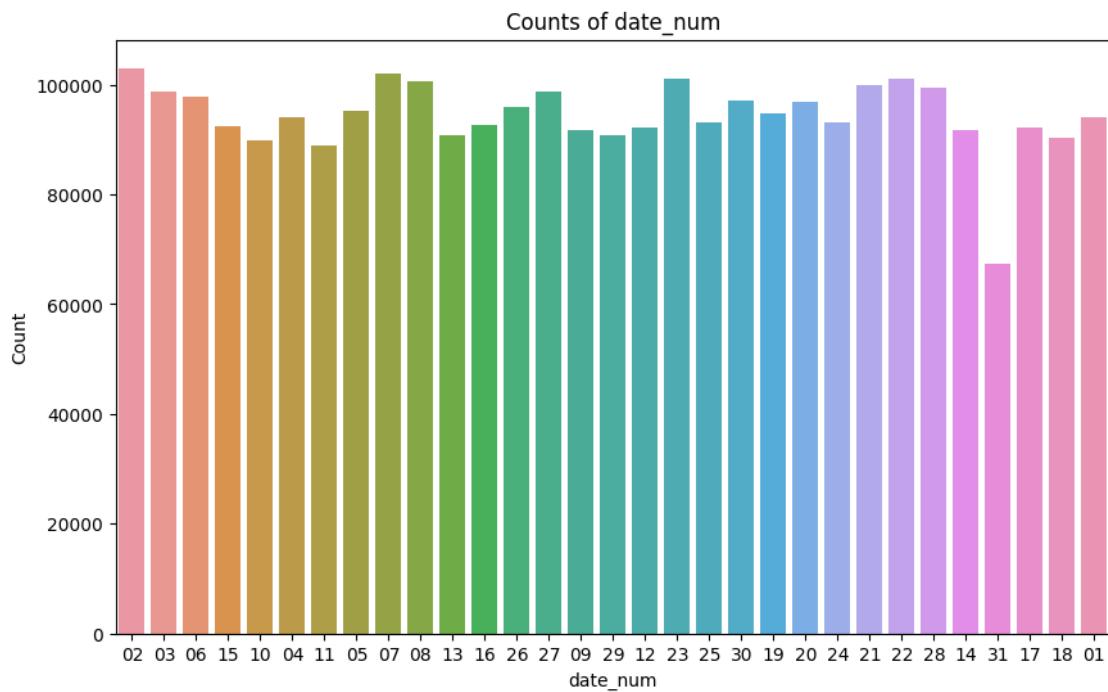
```
[ ]: # univariate analysis
```

```
for column in final_dataset.columns:  
    variable_type = final_dataset[column].dtype  
  
    summary_stats = final_dataset[column].describe()  
  
    plt.figure(figsize=(10, 6))  
  
    # For numerical variables, create a histogram  
    if variable_type in ['int64', 'float64']:  
        sns.histplot(data=final_dataset, x=column, kde=True)  
        plt.title(f'Distribution of {column}')  
        plt.xlabel(column)  
        plt.ylabel('Frequency')  
  
    # For categorical variables, create a bar plot  
    else:  
        sns.countplot(data=final_dataset, x=column)  
        plt.title(f'Counts of {column}')  
        plt.xlabel(column)  
        plt.ylabel('Count')  
  
    plt.show()  
  
    # Print summary statistics  
    print(f"Summary Statistics for {column}:")  
    print(summary_stats)
```



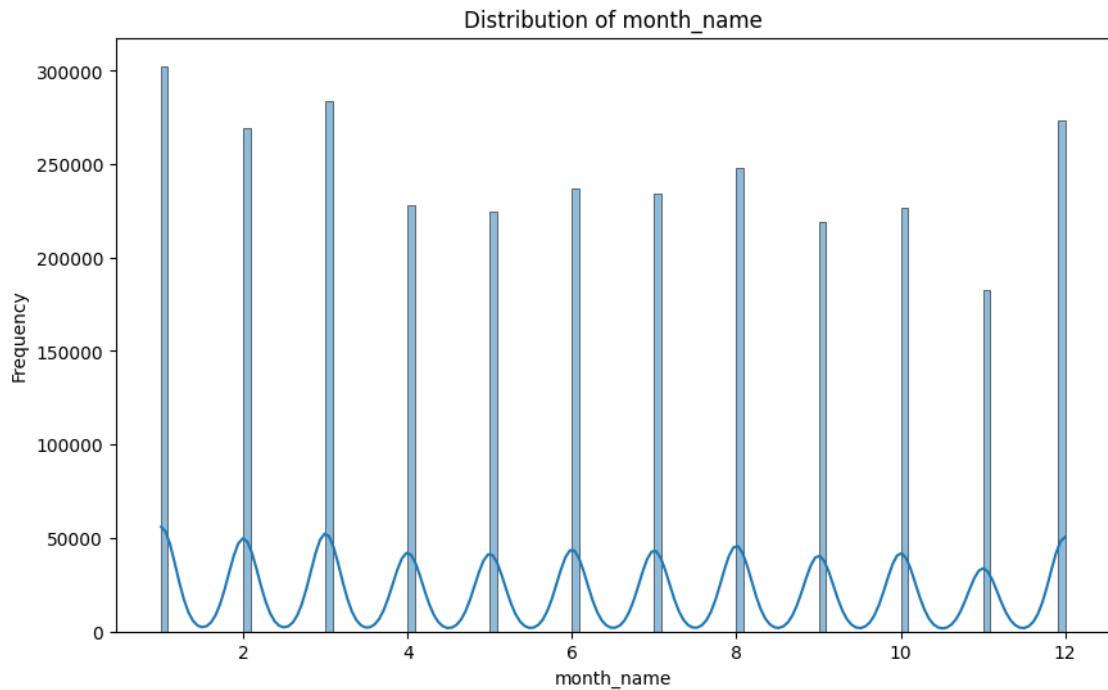
Summary Statistics for date:

```
count      2928483
unique     1034
top       28.12.2013
freq      9415
Name: date, dtype: object
```



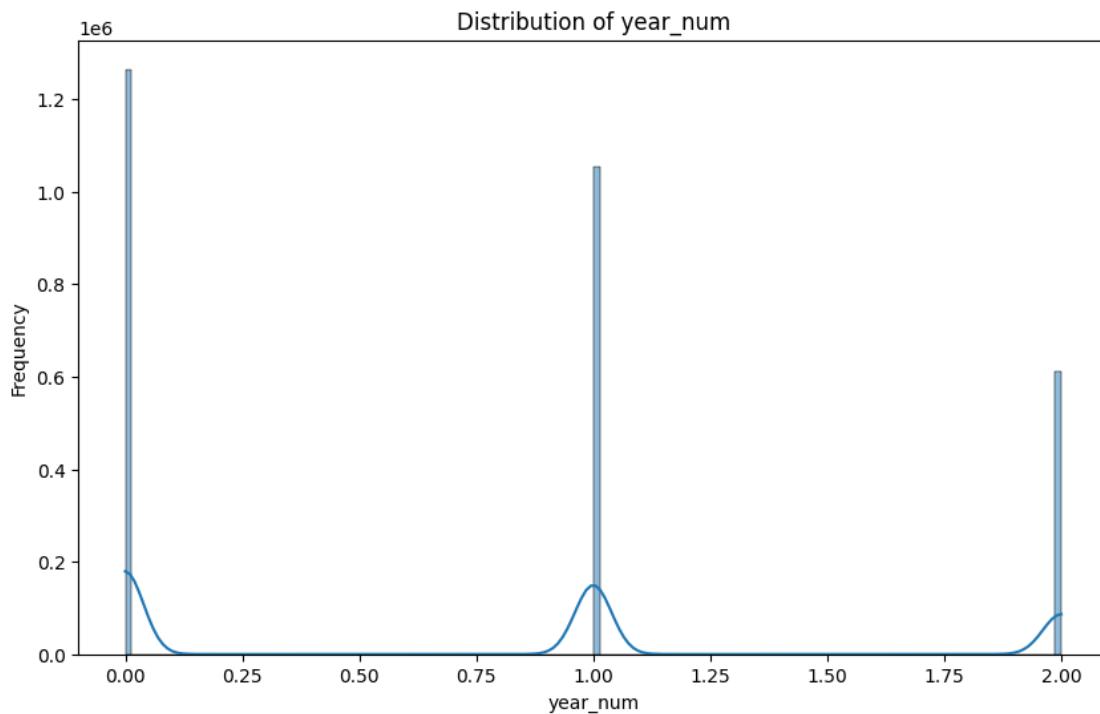
Summary Statistics for date_num:

```
count      2928483
unique       31
top         02
freq      103081
Name: date_num, dtype: object
```



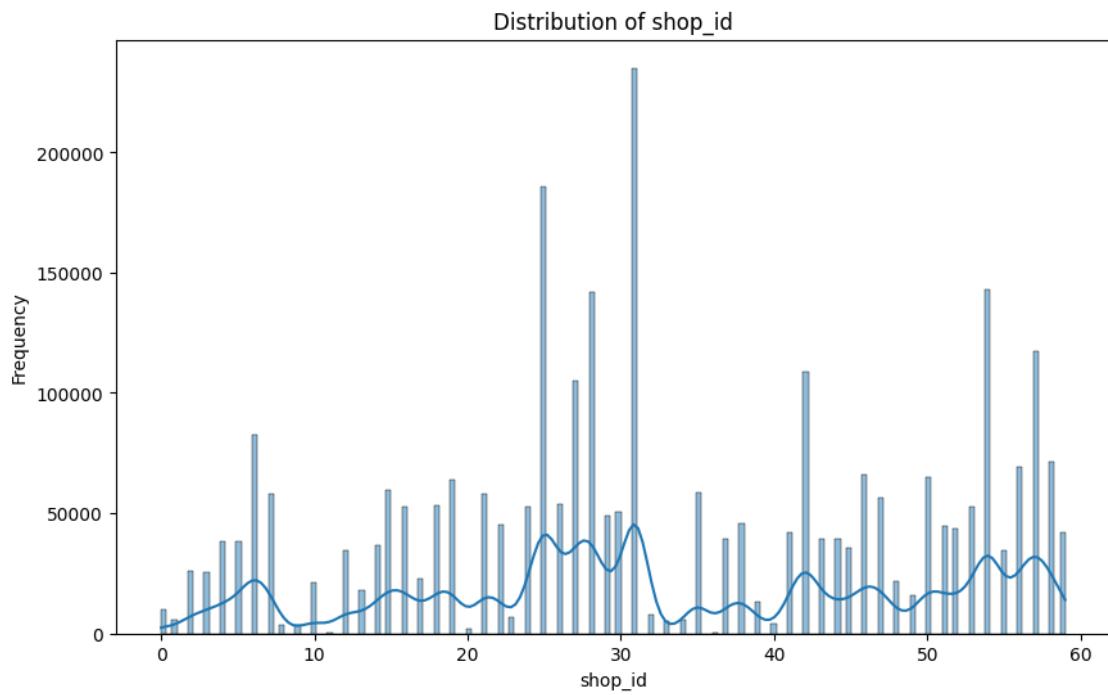
Summary Statistics for month_name:

```
count    2.928483e+06
mean     6.248408e+00
std      3.535921e+00
min      1.000000e+00
25%      3.000000e+00
50%      6.000000e+00
75%      9.000000e+00
max      1.200000e+01
Name: month_name, dtype: float64
```



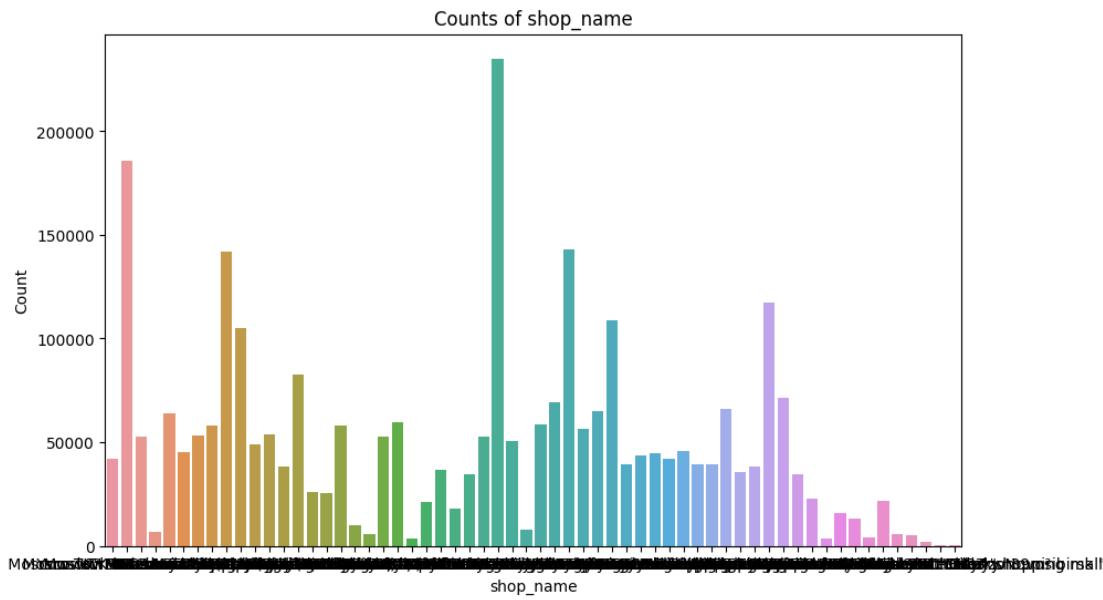
Summary Statistics for year_num:

```
count    2.928483e+06
mean     7.767790e-01
std      7.684598e-01
min      0.000000e+00
25%     0.000000e+00
50%     1.000000e+00
75%     1.000000e+00
max      2.000000e+00
Name: year_num, dtype: float64
```



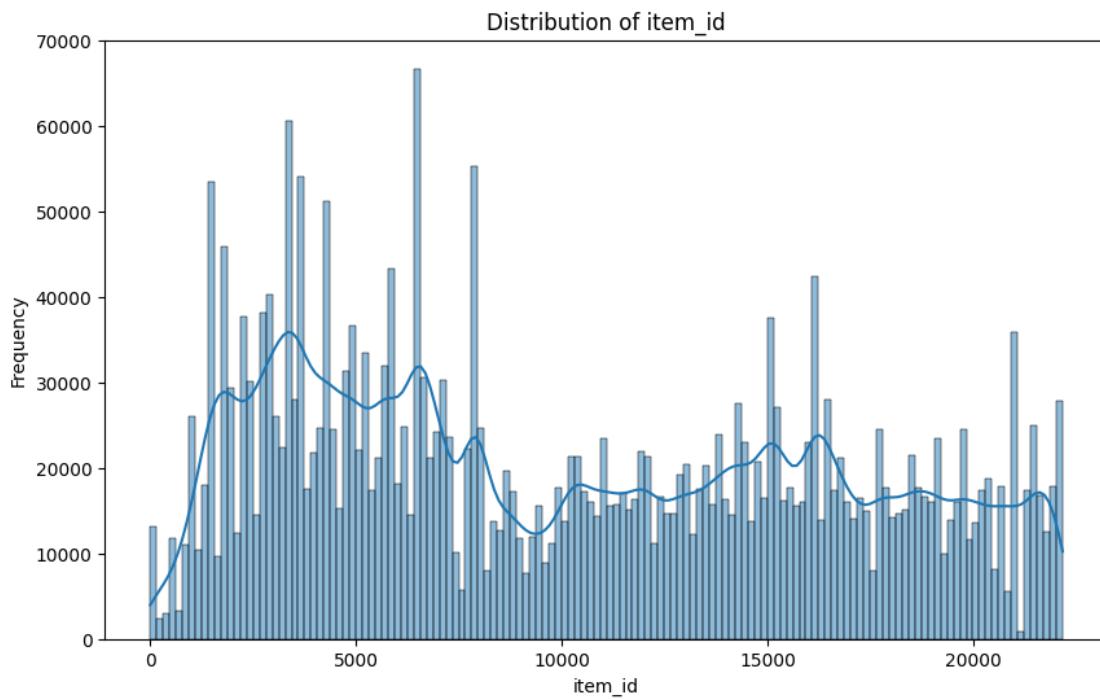
Summary Statistics for shop_id:

```
count    2.928483e+06
mean     3.300296e+01
std      1.622543e+01
min      0.000000e+00
25%      2.200000e+01
50%      3.100000e+01
75%      4.700000e+01
max      5.900000e+01
Name: shop_id, dtype: float64
```



Summary Statistics for shop_name:

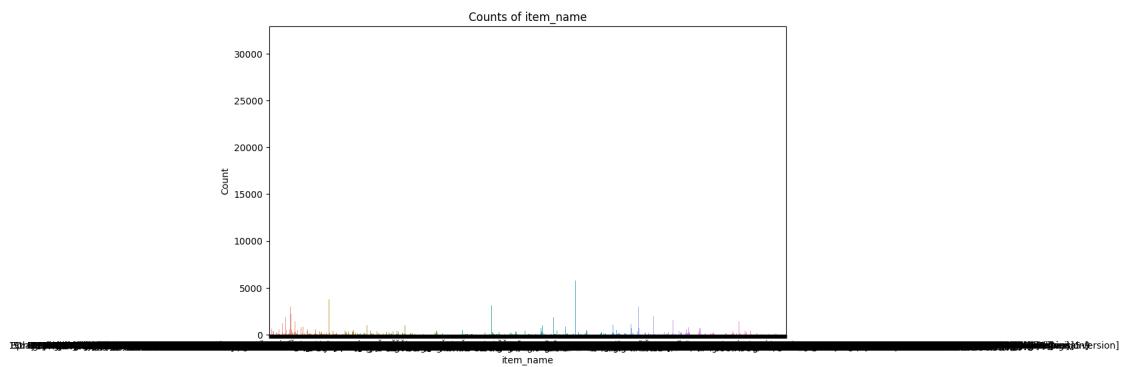
```
count          2928483
unique         60
top      "Moscow TC" "Semenovsky"
freq          235185
Name: shop_name, dtype: object
```



```
Summary Statistics for item_id:
```

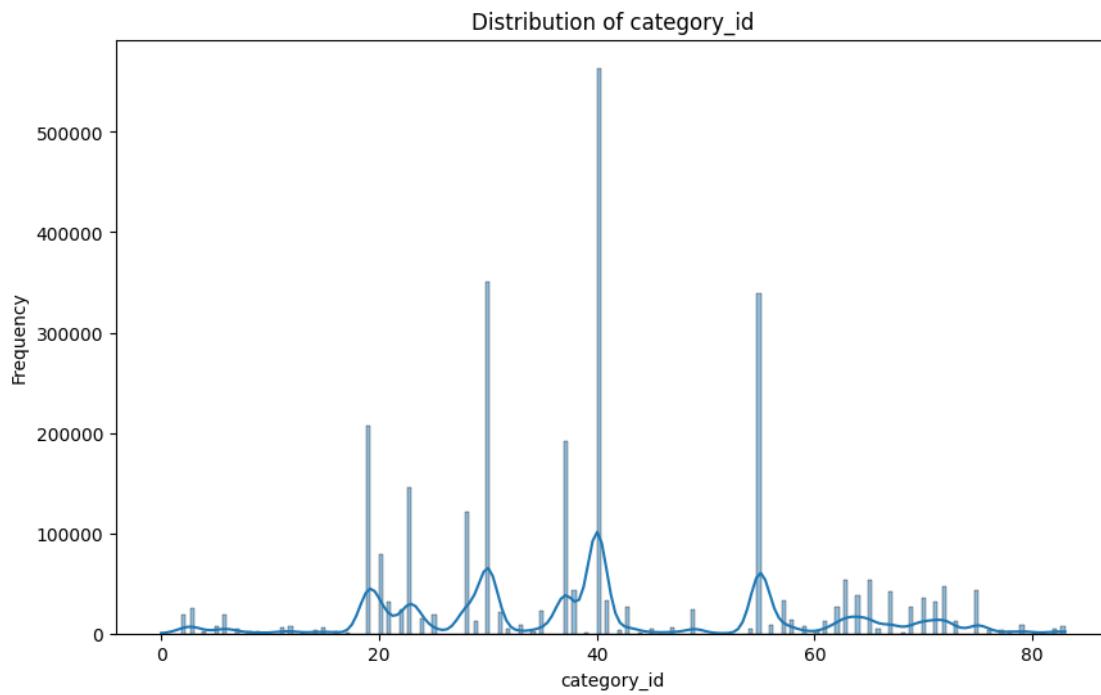
```
count      2.928483e+06  
mean      1.020028e+04  
std       6.324391e+03  
min       0.000000e+00  
25%       4.477000e+03  
50%       9.355000e+03  
75%       1.569100e+04  
max       2.216900e+04
```

```
Name: item_id, dtype: float64
```



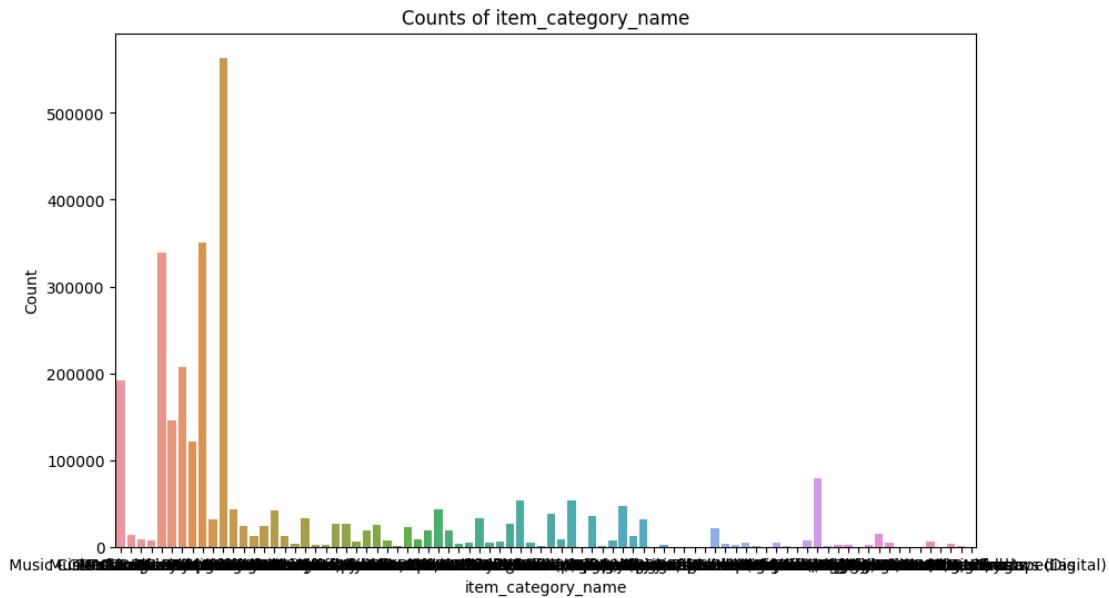
```
Summary Statistics for item_name:
```

```
count                      2928483  
unique                     21782  
top          Corporate package T-shirt 1C Interest white (3...  
freq                      31336  
Name: item_name, dtype: object
```



Summary Statistics for category_id:

```
count    2.928483e+06
mean     4.001637e+01
std      1.709809e+01
min      0.000000e+00
25%      2.800000e+01
50%      4.000000e+01
75%      5.500000e+01
max      8.300000e+01
Name: category_id, dtype: float64
```



Summary Statistics for item_category_name:

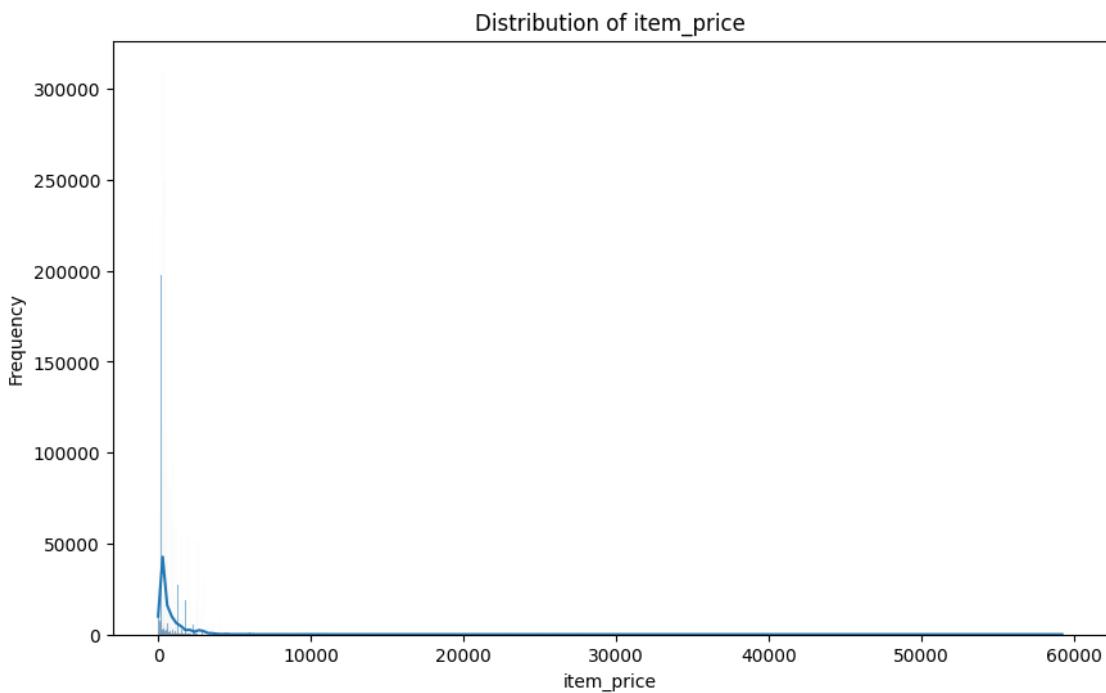
count 2928483

unique 84

top Cinema - DVD

freq 563937

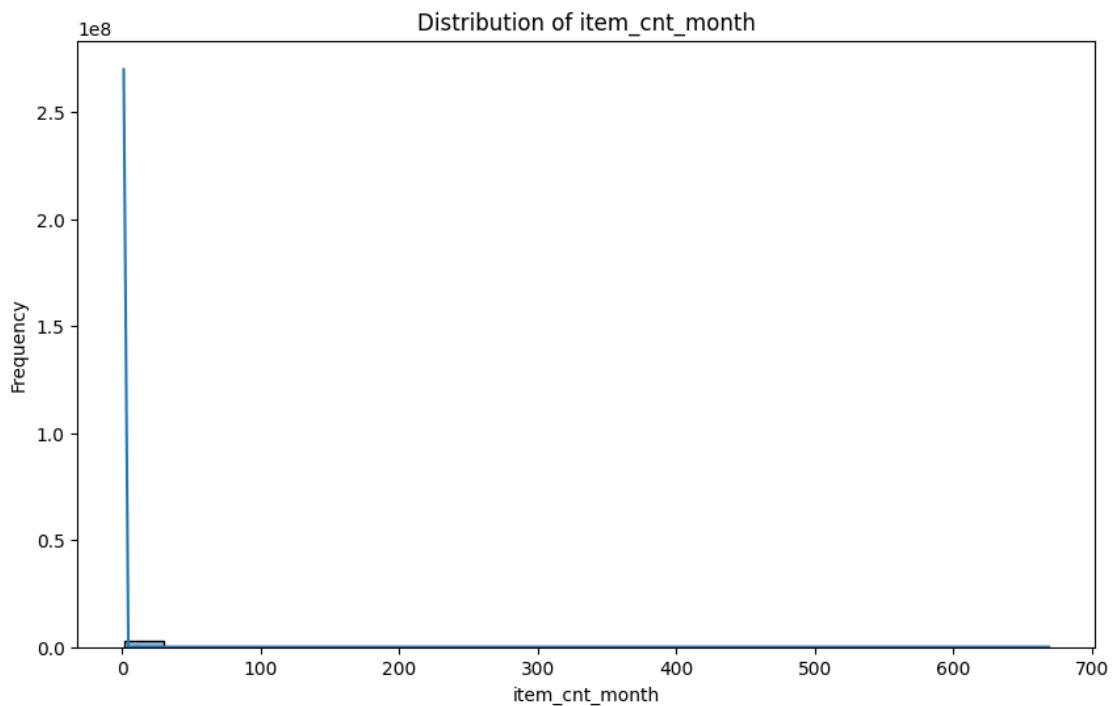
Name: item_category_name, dtype: object



```
Summary Statistics for item_price:
```

```
count    2.928483e+06
mean     8.893627e+02
std      1.718155e+03
min      7.000000e-02
25%     2.490000e+02
50%     3.990000e+02
75%     9.990000e+02
max     5.920000e+04
```

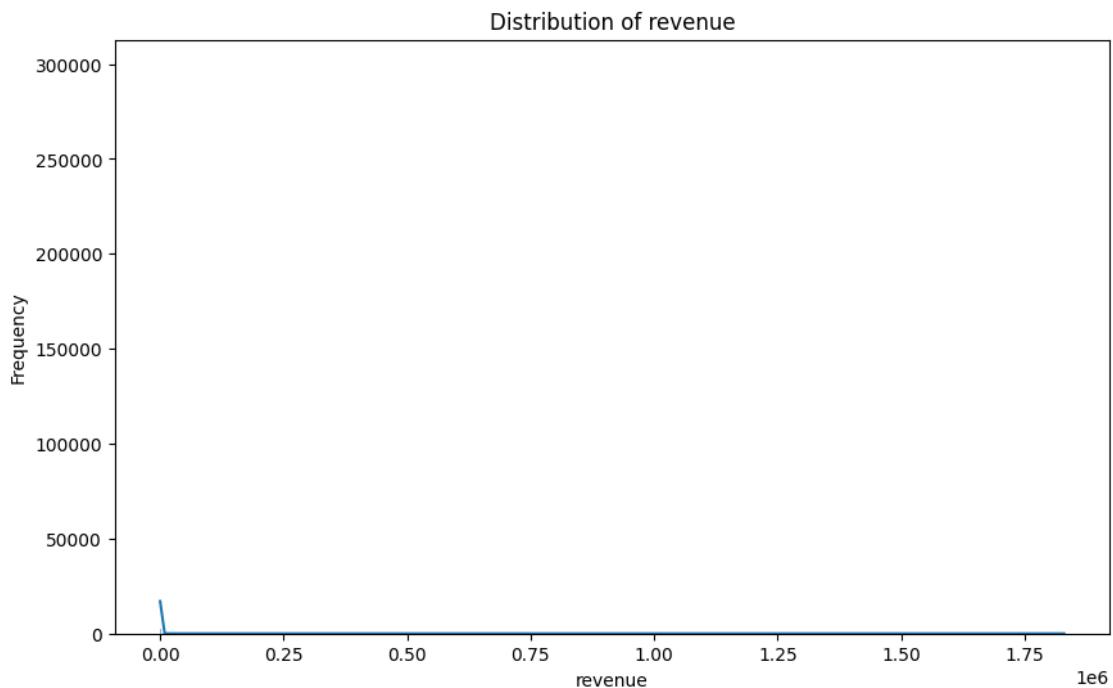
```
Name: item_price, dtype: float64
```



```
Summary Statistics for item_cnt_month:
```

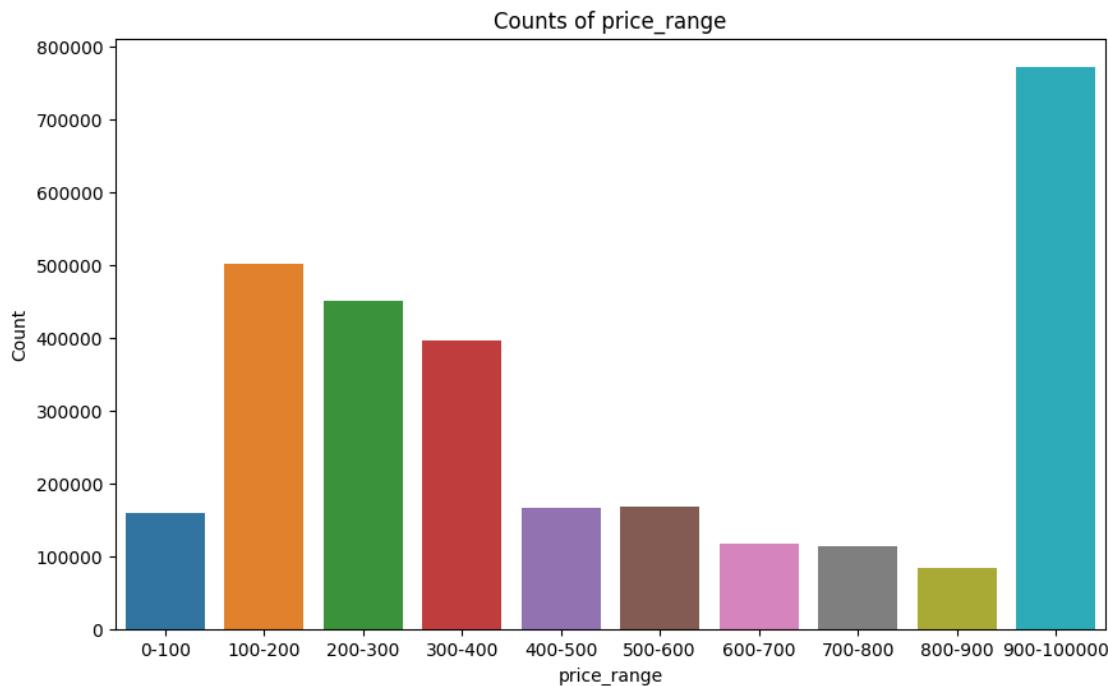
```
count    2.928483e+06
mean     1.247257e+00
std      2.217429e+00
min      1.000000e+00
25%     1.000000e+00
50%     1.000000e+00
75%     1.000000e+00
max     6.690000e+02
```

```
Name: item_cnt_month, dtype: float64
```



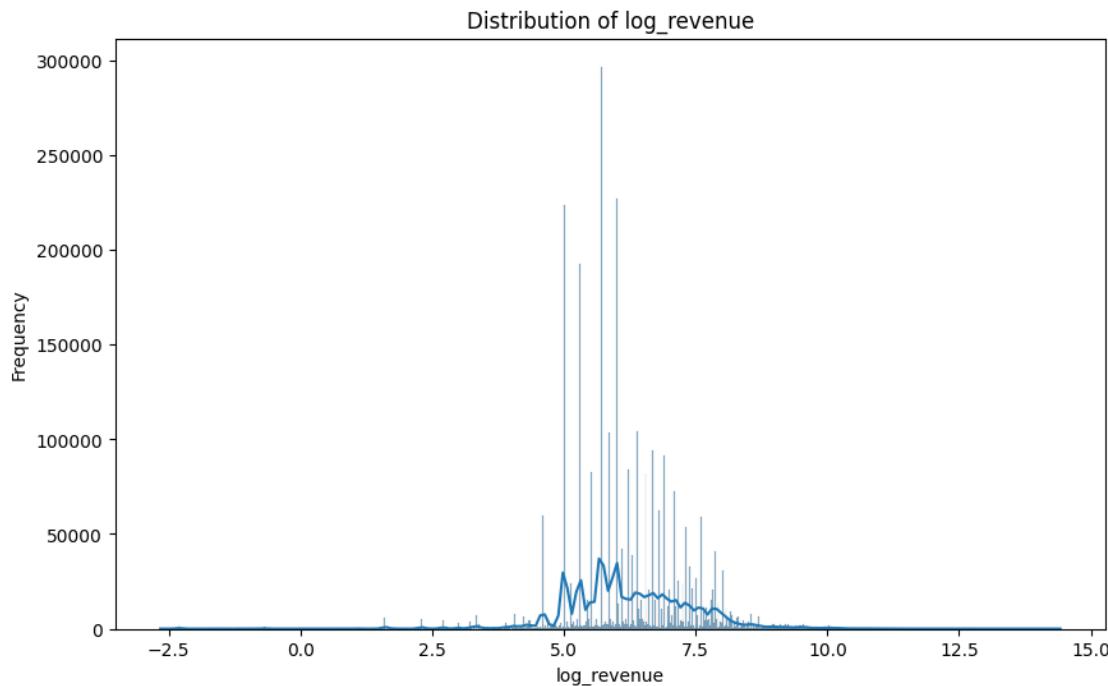
Summary Statistics for revenue:

```
count    2.928483e+06
mean     1.164267e+03
std      5.684853e+03
min      7.000000e-02
25%      2.490000e+02
50%      4.490000e+02
75%      1.090000e+03
max      1.829990e+06
Name: revenue, dtype: float64
```



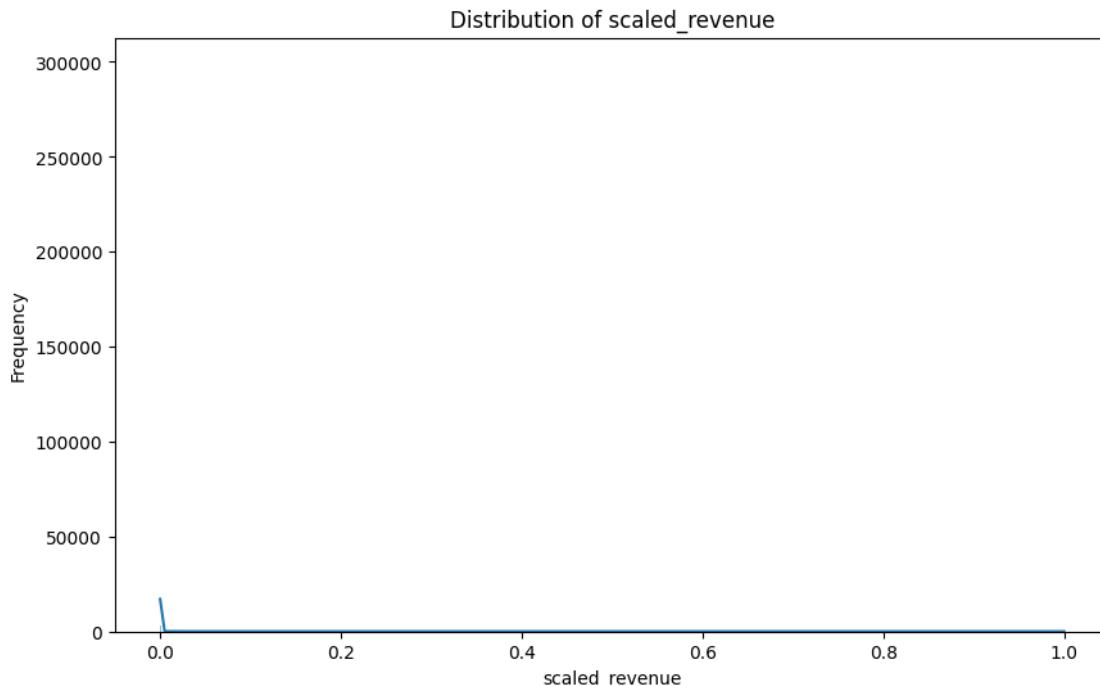
Summary Statistics for price_range:

```
count      2928483
unique       10
top    900-100000
freq     772112
Name: price_range, dtype: object
```



Summary Statistics for log_revenue:

```
count    2.928483e+06
mean     6.254676e+00
std      1.171779e+00
min     -2.659260e+00
25%     5.517453e+00
50%     6.107023e+00
75%     6.993933e+00
max     1.441982e+01
Name: log_revenue, dtype: float64
```



```
Summary Statistics for scaled_revenue:
count      2.928483e+06
mean       6.361770e-04
std        3.106494e-03
min        0.000000e+00
25%        1.360281e-04
50%        2.453183e-04
75%        5.955934e-04
max        1.000000e+00
Name: scaled_revenue, dtype: float64
```

```
[ ]: #bivariate analysis

#can analysis by changing var1 and var2
var1 = 'item_price'
var2 = 'item_cnt_month'

var1_type = final_dataset[var1].dtype
var2_type = final_dataset[var2].dtype

# Scatter Plot for Numerical vs. Numerical
if var1_type in ['int64', 'float64'] and var2_type in ['int64', 'float64']:
    plt.figure(figsize=(10, 6))
    sns.scatterplot(data=final_dataset, x=var1, y=var2)
    plt.title(f'Scatter Plot: {var1} vs. {var2}')
```

```

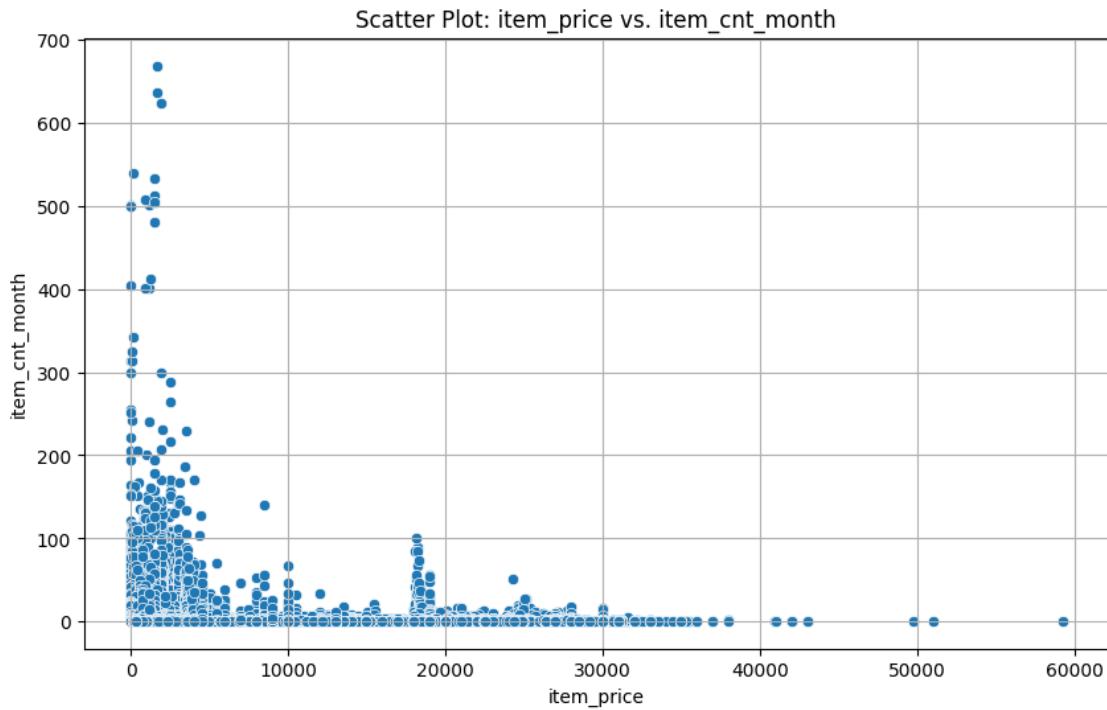
plt.xlabel(var1)
plt.ylabel(var2)
plt.grid(True)
plt.show()

# Box Plot for Categorical vs. Numerical
elif var1_type in ['object', 'category'] and var2_type in ['int64', 'float64']:
    plt.figure(figsize=(10, 6))
    sns.boxplot(data=final_dataset, x=var1, y=var2)
    plt.title(f'Box Plot: {var1} vs. {var2}')
    plt.xlabel(var1)
    plt.ylabel(var2)
    plt.grid(True)
    plt.show()

# Bar Plot for Categorical vs. Categorical
elif var1_type in ['object', 'category'] and var2_type in ['object', 'category']:
    crosstab = pd.crosstab(final_dataset[var1], final_dataset[var2])
    crosstab.plot(kind='bar', stacked=True, figsize=(10, 6))
    plt.title(f'Bar Plot: {var1} vs. {var2}')
    plt.xlabel(var1)
    plt.ylabel('Count')
    plt.grid(True)
    plt.show()

# Print correlation for Numerical vs. Numerical
if var1_type in ['int64', 'float64'] and var2_type in ['int64', 'float64']:
    correlation = final_dataset[[var1, var2]].corr().iloc[0, 1]
    print(f'Correlation between {var1} and {var2}: {correlation:.2f}')

```



Correlation between item_price and item_cnt_month: 0.01

```
[ ]: #Exploratory Data Analysis (EDA)

print("Dataset Overview:")
print(final_dataset.info())

print("\nSummary Statistics for Numerical Variables:")
print(final_dataset.describe())

print("\nMissing Values:")
print(final_dataset.isnull().sum())

numerical_columns = ['month_name', 'year_num', 'shop_id', 'item_id',
                     'category_id', 'item_price', 'item_cnt_month', 'revenue', 'log_revenue',
                     'scaled_revenue']

for column in numerical_columns:
    plt.figure(figsize=(8, 4))
    sns.histplot(data=final_dataset, x=column, kde=True, bins=20)
    plt.title(f'Distribution of {column}')
    plt.xlabel(column)
    plt.ylabel('Frequency')
    plt.show()
```

```

# Visualize relationships between variables with a correlation matrix for numerical variables
correlation_matrix = final_dataset[numerical_columns].corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap for Numerical Variables")
plt.show()

# Explore categorical variables with bar plots
categorical_columns = ['shop_name', 'item_name', 'item_category_name', 'price_range']

for column in categorical_columns:
    plt.figure(figsize=(10, 6))
    sns.countplot(data=final_dataset, x=column)
    plt.title(f'Counts of {column}')
    plt.xlabel(column)
    plt.ylabel('Count')
    plt.xticks(rotation=90)
    plt.show()

```

Dataset Overview:

```

<class 'pandas.core.frame.DataFrame'>
Index: 2928483 entries, 0 to 2935848
Data columns (total 16 columns):
 #   Column           Dtype  
 ---  --  
 0   date             object 
 1   date_num         object 
 2   month_name       int64  
 3   year_num         int64  
 4   shop_id          int64  
 5   shop_name        object 
 6   item_id          int64  
 7   item_name        object 
 8   category_id      int64  
 9   item_category_name object 
 10  item_price       float64
 11  item_cnt_month   int64  
 12  revenue          float64
 13  price_range      category
 14  log_revenue      float64
 15  scaled_revenue   float64
dtypes: category(1), float64(4), int64(6), object(5)
memory usage: 360.3+ MB
None

```

Summary Statistics for Numerical Variables:

	month_name	year_num	shop_id	item_id	category_id	\
count	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	
mean	6.248408e+00	7.767790e-01	3.300296e+01	1.020028e+04	4.001637e+01	
std	3.535921e+00	7.684598e-01	1.622543e+01	6.324391e+03	1.709809e+01	
min	1.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	
25%	3.000000e+00	0.000000e+00	2.200000e+01	4.477000e+03	2.800000e+01	
50%	6.000000e+00	1.000000e+00	3.100000e+01	9.355000e+03	4.000000e+01	
75%	9.000000e+00	1.000000e+00	4.700000e+01	1.569100e+04	5.500000e+01	
max	1.200000e+01	2.000000e+00	5.900000e+01	2.216900e+04	8.300000e+01	

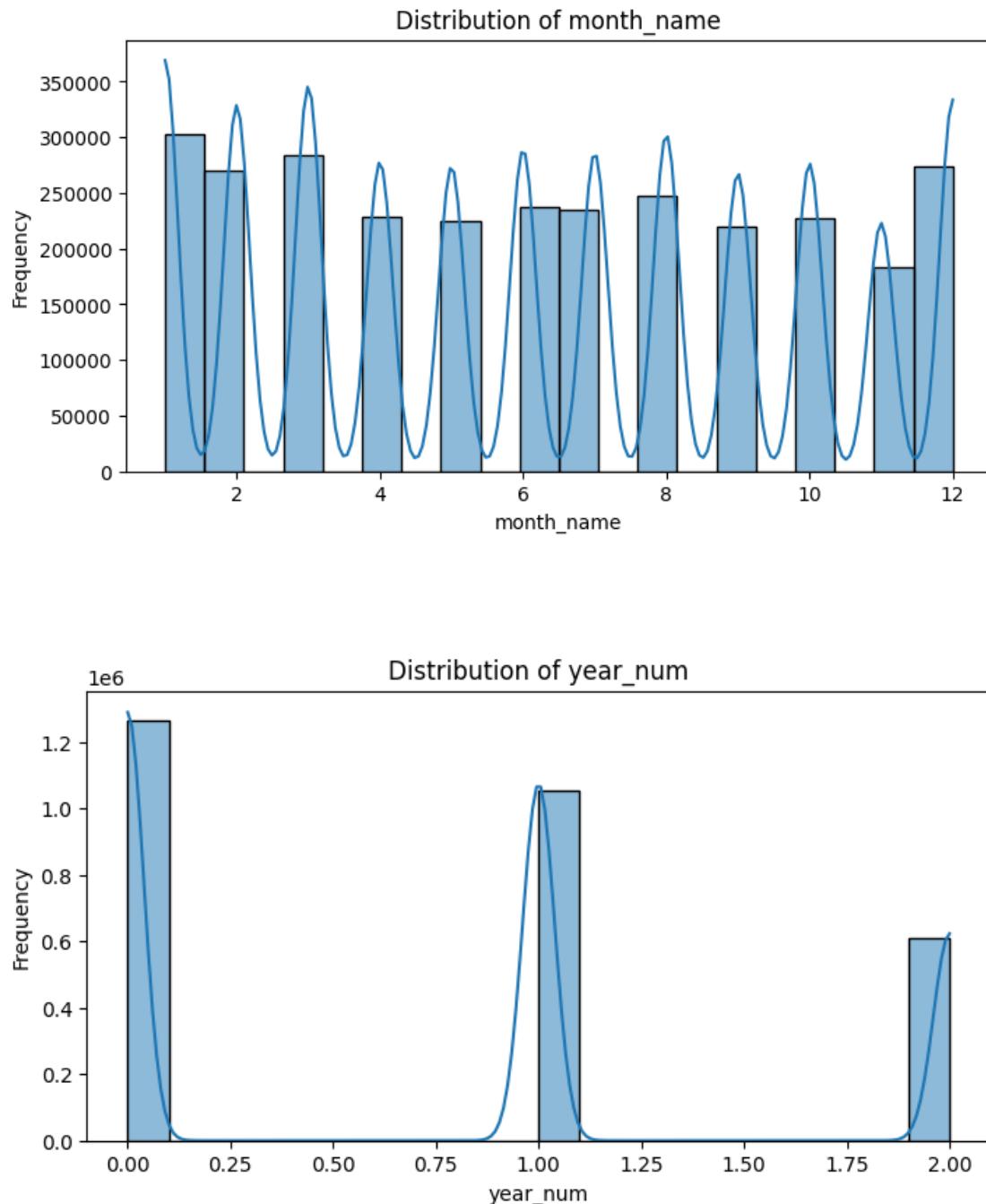
	item_price	item_cnt_month	revenue	log_revenue	\
count	2.928483e+06	2.928483e+06	2.928483e+06	2.928483e+06	
mean	8.893627e+02	1.247257e+00	1.164267e+03	6.254676e+00	
std	1.718155e+03	2.217429e+00	5.684853e+03	1.171779e+00	
min	7.000000e-02	1.000000e+00	7.000000e-02	-2.659260e+00	
25%	2.490000e+02	1.000000e+00	2.490000e+02	5.517453e+00	
50%	3.990000e+02	1.000000e+00	4.490000e+02	6.107023e+00	
75%	9.990000e+02	1.000000e+00	1.090000e+03	6.993933e+00	
max	5.920000e+04	6.690000e+02	1.829990e+06	1.441982e+01	

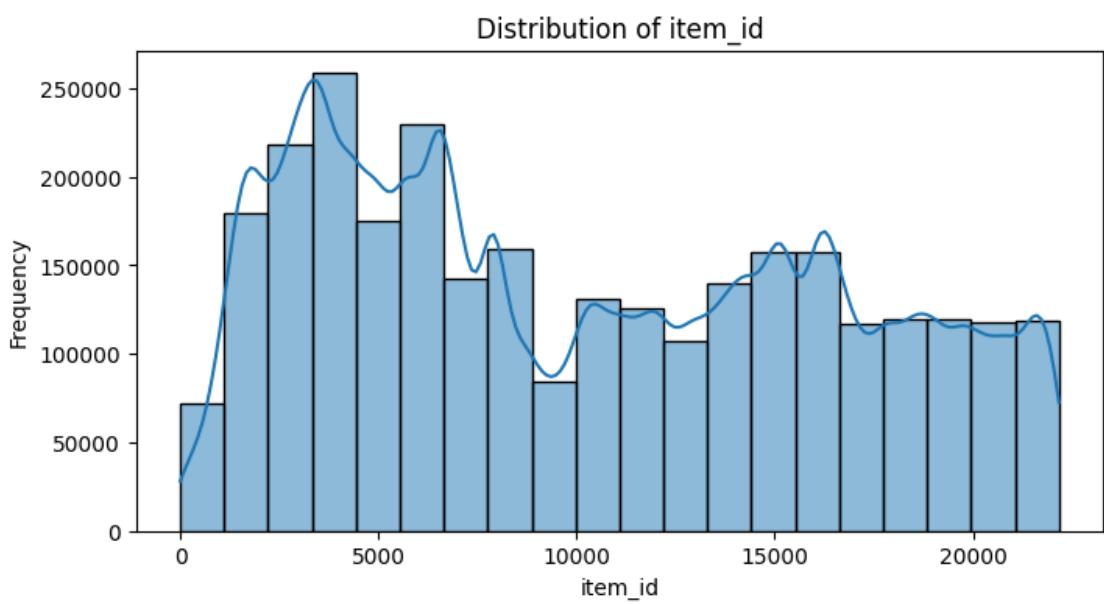
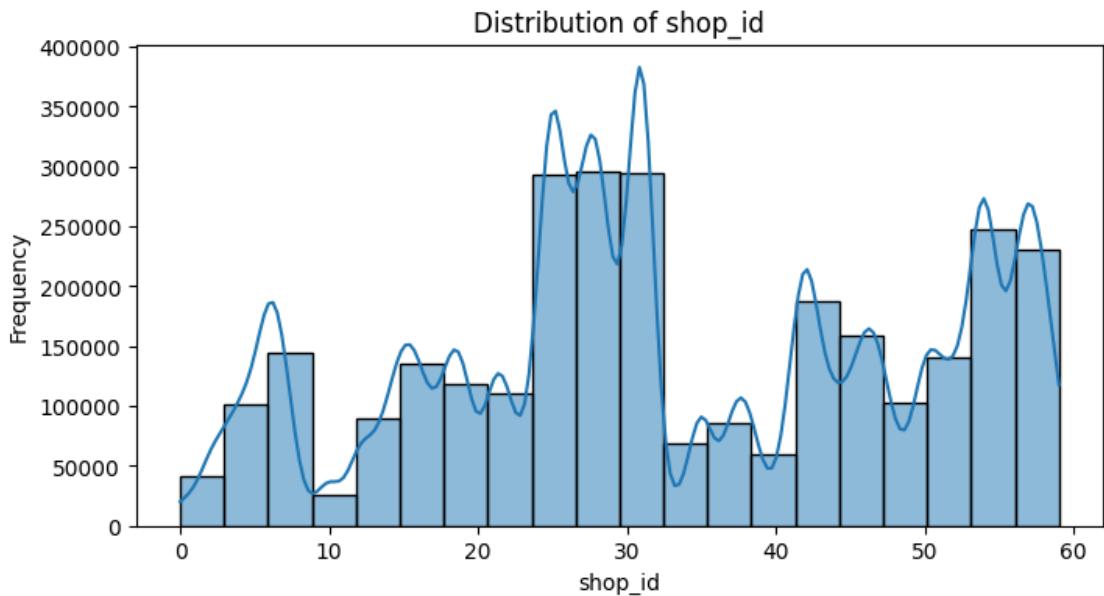
	scaled_revenue
count	2.928483e+06
mean	6.361770e-04
std	3.106494e-03
min	0.000000e+00
25%	1.360281e-04
50%	2.453183e-04
75%	5.955934e-04
max	1.000000e+00

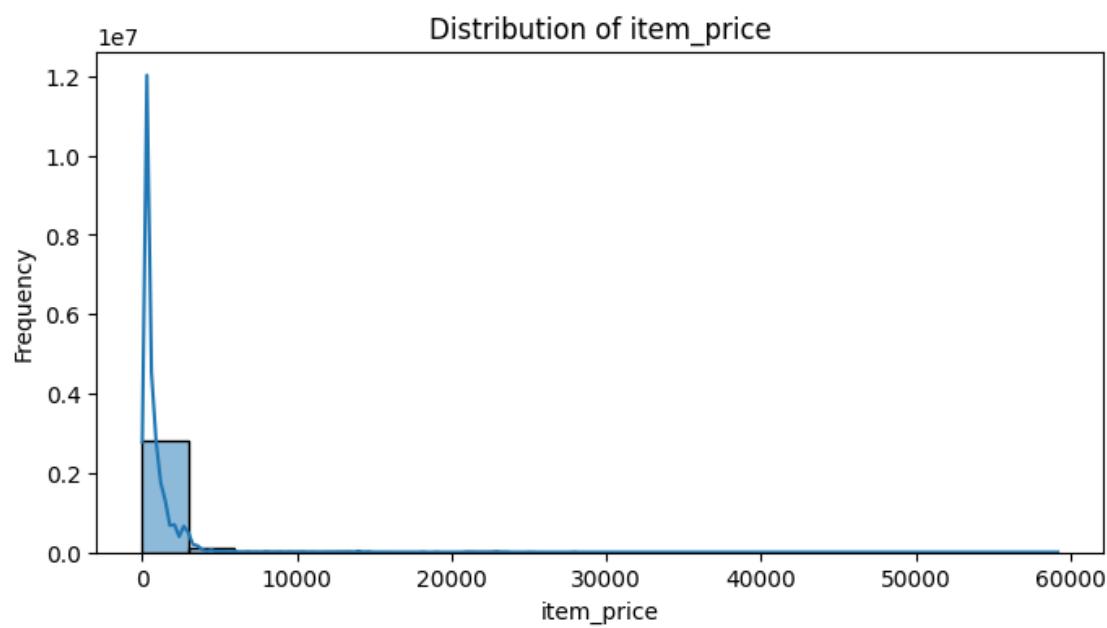
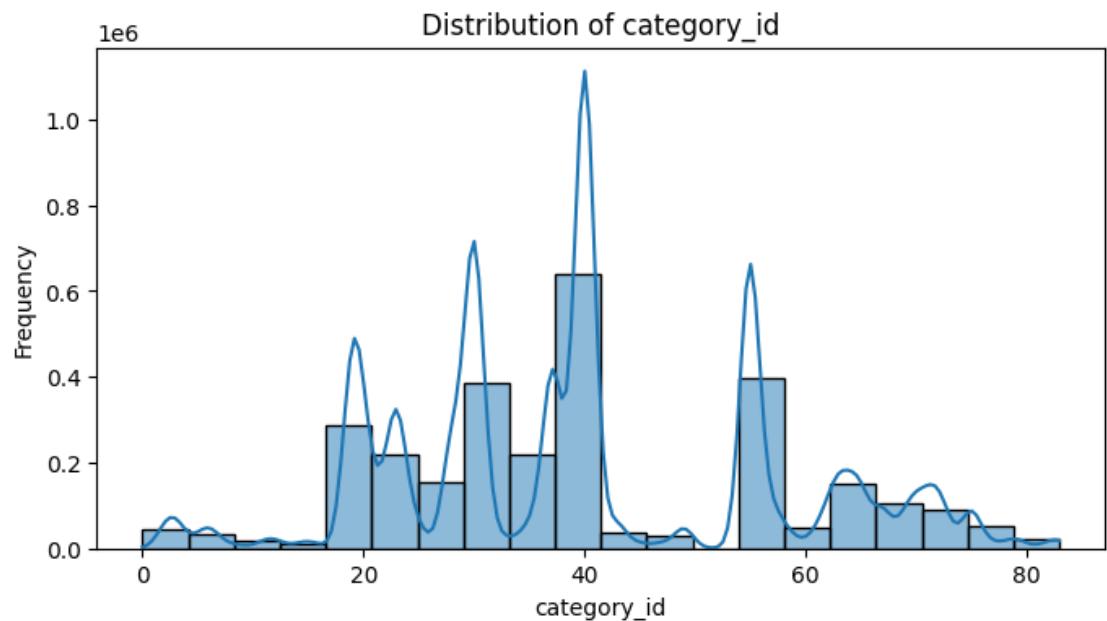
Missing Values:

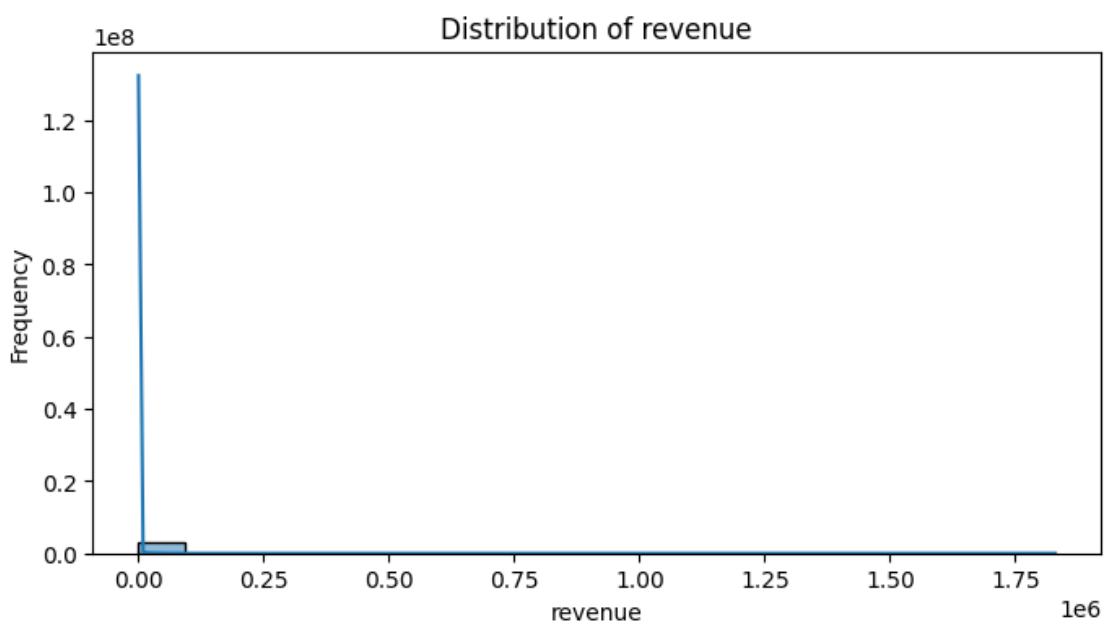
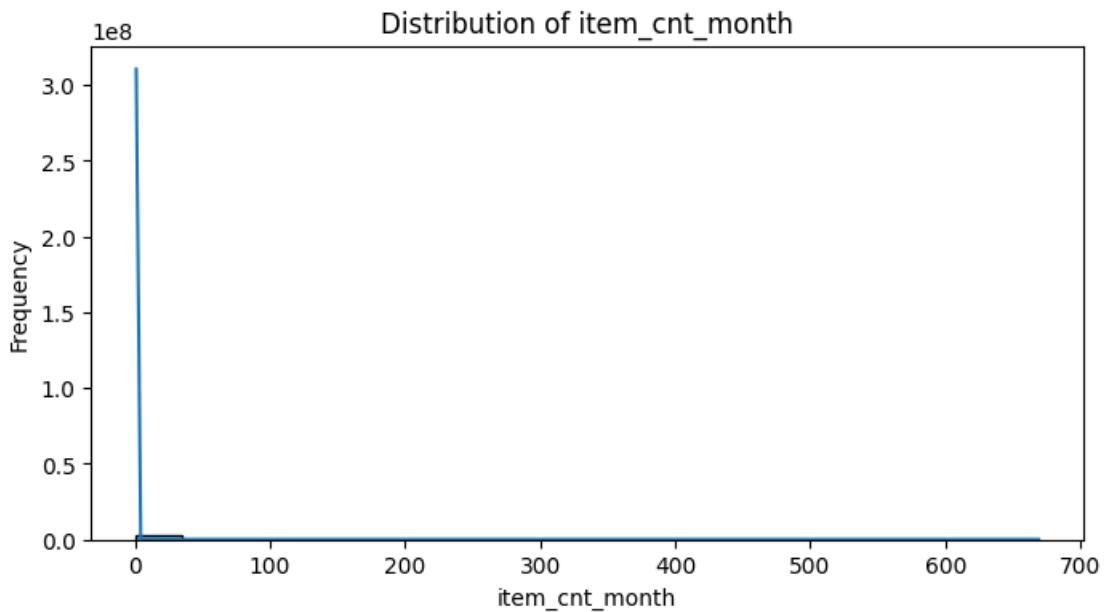
date	0
date_num	0
month_name	0
year_num	0
shop_id	0
shop_name	0
item_id	0
item_name	0
category_id	0
item_category_name	0
item_price	0
item_cnt_month	0
revenue	0
price_range	0
log_revenue	0

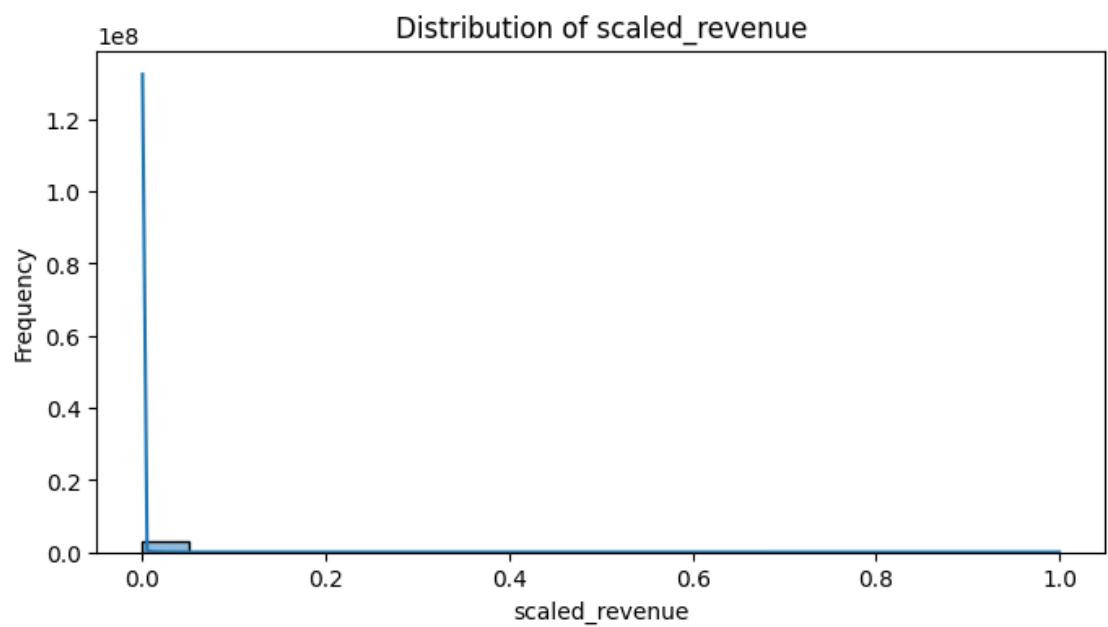
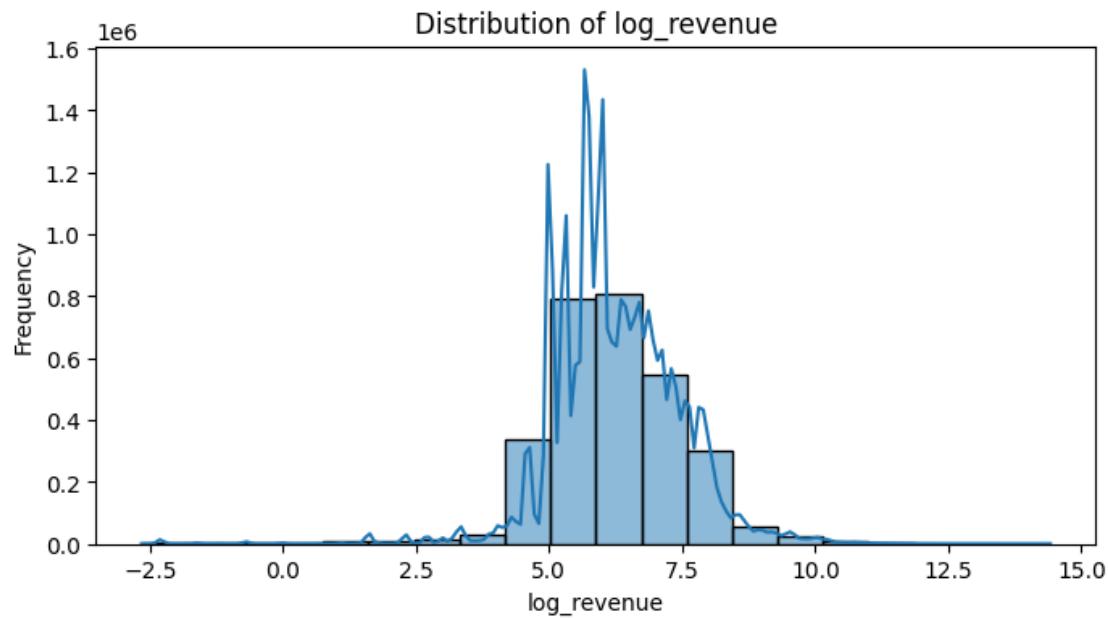
```
scaled_revenue      0  
dtype: int64
```

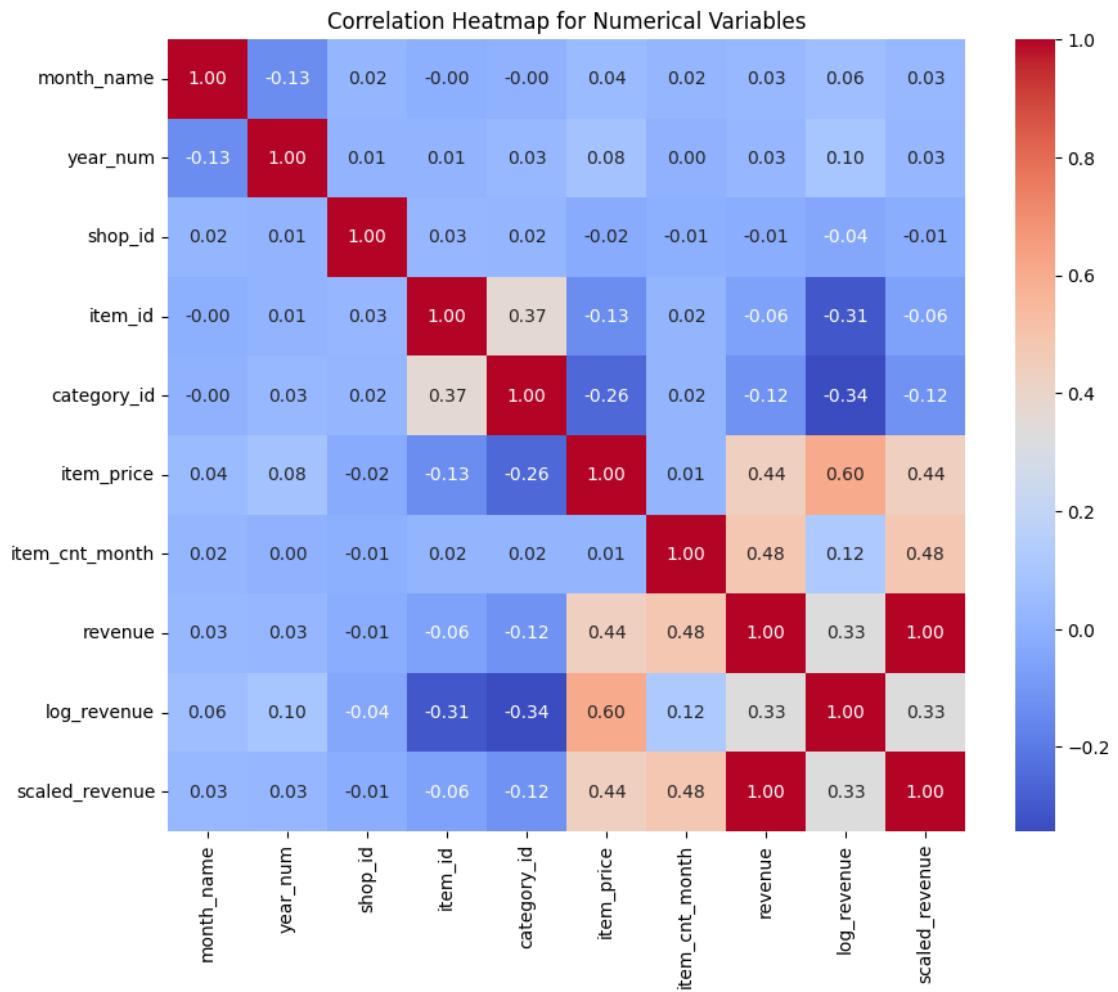




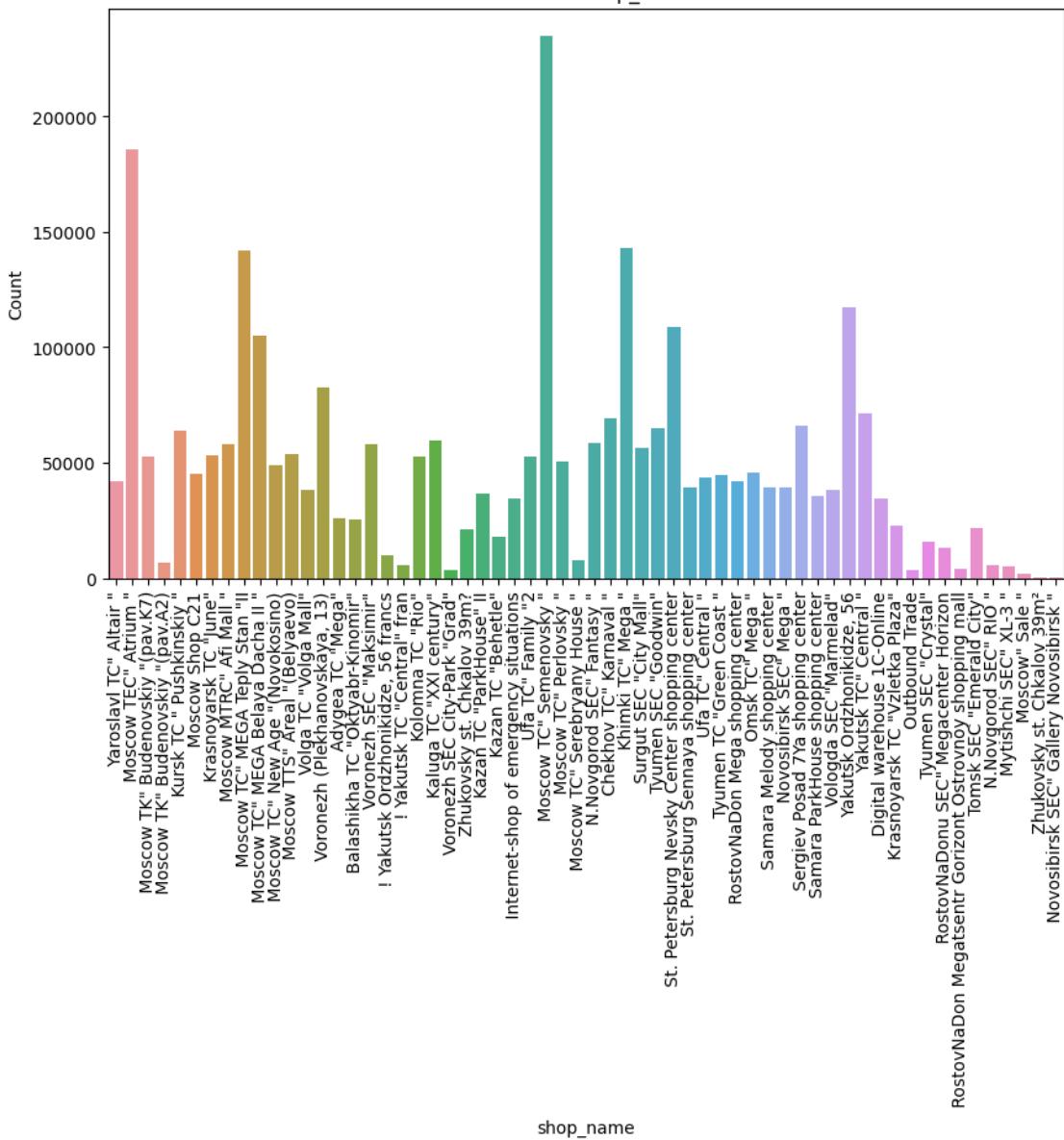


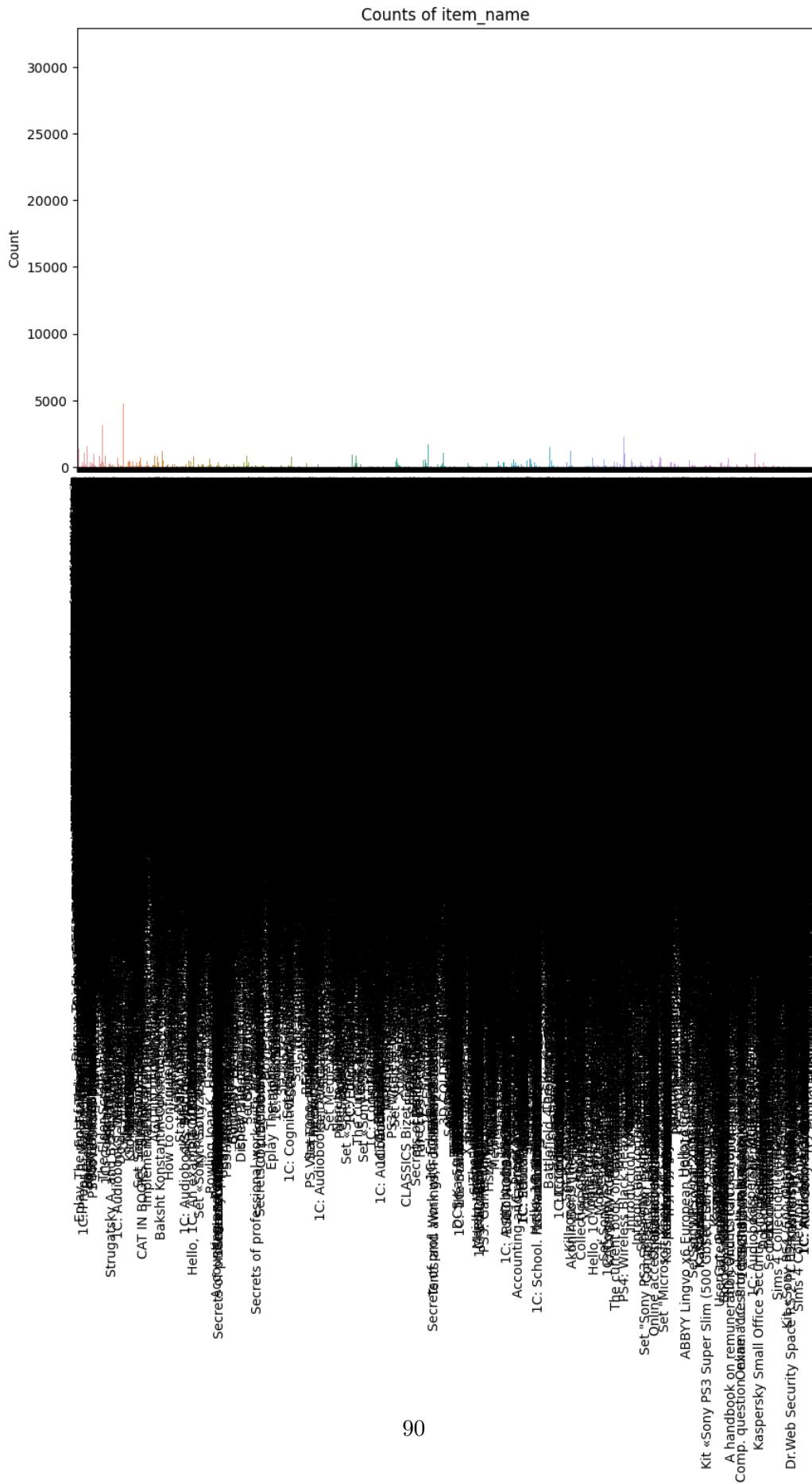


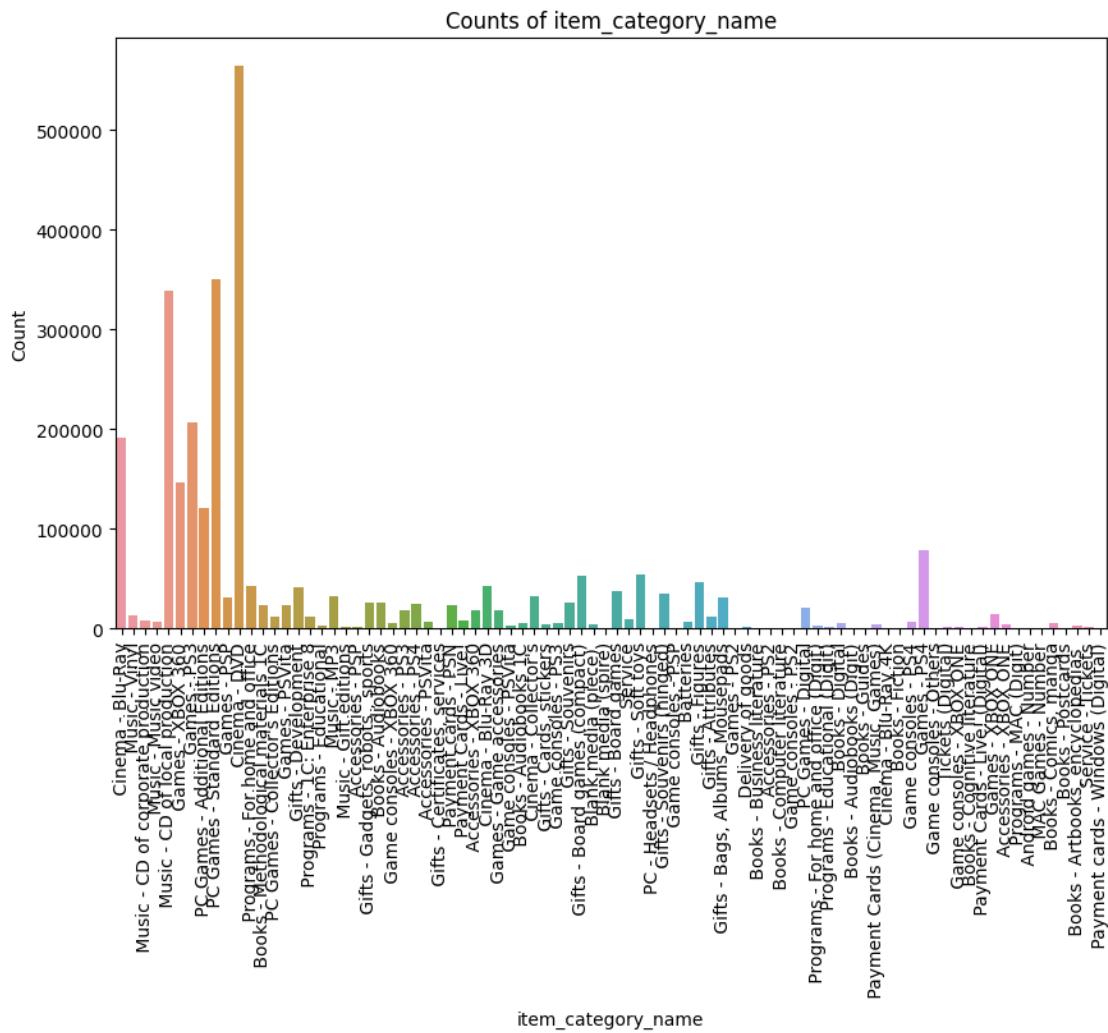


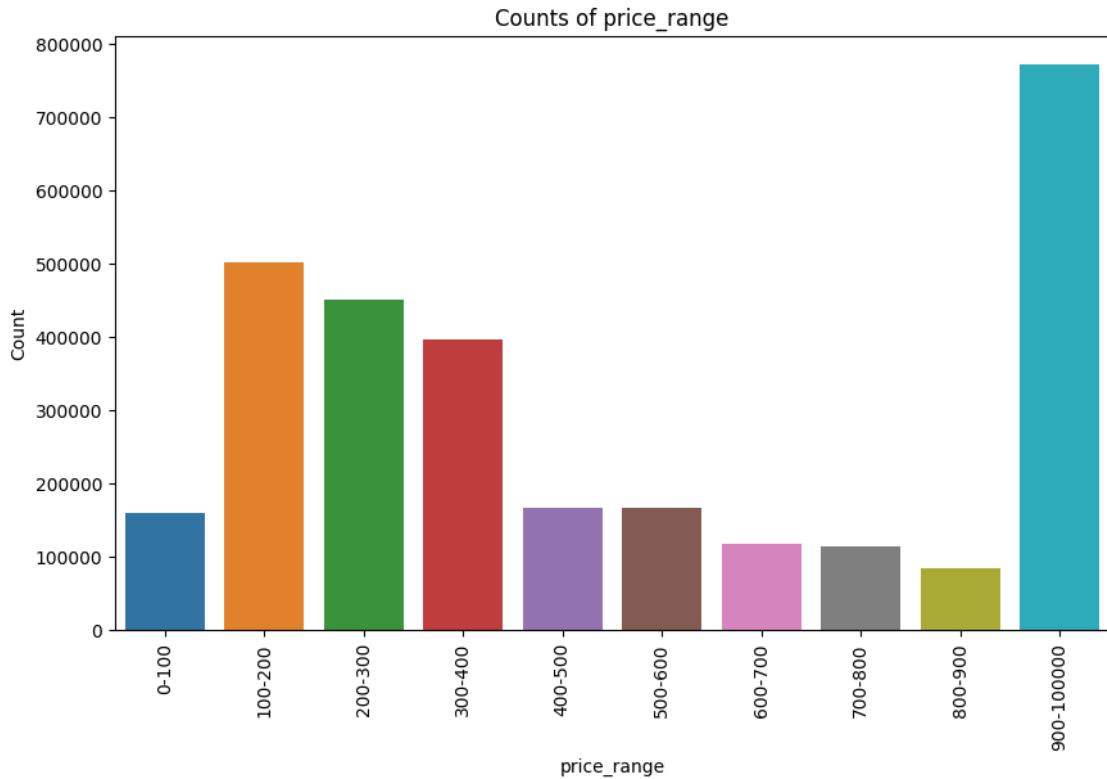


Counts of shop_name









```
[ ]: #inferential analysis

np.random.seed(42)
data = np.random.normal(loc=70, scale=10, size=100)

# Create a DataFrame from the generated data
df = pd.DataFrame({'measurement': data})

# Calculate the sample mean and standard deviation
sample_mean = df['measurement'].mean()
sample_std = df['measurement'].std()

# Define a hypothetical population mean for comparison
population_mean = 75

# Perform a t-test to compare the sample mean with the population mean
t_statistic, p_value = stats.ttest_1samp(df['measurement'], population_mean)

# Print results
print(f"Sample Mean: {sample_mean:.2f}")
print(f"Sample Standard Deviation: {sample_std:.2f}")
print(f"Population Mean: {population_mean}")
```

```

print(f"T-Statistic: {t_statistic:.2f}")
print(f"P-Value: {p_value:.4f}")

# Determine statistical significance
alpha = 0.05 # Significance level (adjust as needed)
if p_value < alpha:
    print("Reject the null hypothesis: The sample mean is statistically different from the population mean.")
else:
    print("Fail to reject the null hypothesis: There is no significant difference between the sample mean and the population mean.")

```

Sample Mean: 68.96
 Sample Standard Deviation: 9.08
 Population Mean: 75
 T-Statistic: -6.65
 P-Value: 0.0000
 Reject the null hypothesis: The sample mean is statistically different from the population mean.

```

[ ]: #diagnostic analytics

# Generate a hypothetical dataset
np.random.seed(42)
X = np.random.rand(100, 1) * 10
y = 3 * X + 2 + np.random.randn(100, 1)

# Create a DataFrame from the generated data
df = pd.DataFrame({'X': X.flatten(), 'y': y.flatten()})

# Diagnostic Plots
plt.figure(figsize=(12, 6))

plt.tight_layout()
plt.show()

```

<Figure size 1200x600 with 0 Axes>

```

[ ]: #qualitative analytics

category_counts = final_dataset['item_category_name'].value_counts()
print(category_counts)

cross_tab = pd.crosstab(final_dataset['shop_name'], final_dataset['item_category_name'])
print(cross_tab)

category_frequency = (final_dataset['price_range'] == 'Low').sum()

```

```

print(f"Frequency of 'Low' price range: {category_frequency}")

average_price_per_category = final_dataset.
    ↪groupby('item_category_name')['item_price'].mean()
print(average_price_per_category)

category_counts.plot(kind='bar', figsize=(10, 6))
plt.title('Item Category Counts')
plt.xlabel('Category')
plt.ylabel('Count')
plt.xticks(rotation=90)
plt.show()

```

item_category_name	
Cinema - DVD	563937
PC Games - Standard Editions	350787
Music - CD of local production	339127
Games - PS3	207371
Cinema - Blu-Ray	191931
	..
Books - Guides	3
Accessories - PS2	2
Books - Postcards	2
Books - Cognitive literature	1
Game consoles - PS2	1
Name: count, Length: 84, dtype: int64	
item_category_name	Accessories - PS2 \
shop_name	
Adygea TC "Mega"	0
Balashikha TC "Oktyabr-Kinomir"	0
Chekhov TC" Karnaval "	0
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	0
Kaluga TC "XXI century"	0
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	0
Khimki TC" Mega "	0
Kolomna TC "Rio"	0
Krasnoyarsk TC "June"	0
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	0
Moscow MTRC" Afi Mall "	0
Moscow Shop C21	0
Moscow TC" MEGA Belaya Dacha II "	0
Moscow TC" MEGA Teply Stan "II	0
Moscow TC" New Age "(Novokosino)	0
Moscow TC" Perlovsky "	0
Moscow TC" Semenovsky "	0

Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	0
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	0
Moscow TTS" Areal "(Belyaev)	0
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	0
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	0
Omsk TC" Mega "	0
Outbound Trade	0
RostovNaDon Mega shopping center	0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	0
Samara Melody shopping center	0
Samara ParkHouse shopping center	1
Sergiev Posad 7Ya shopping center	0
St. Petersburg Nevsky Center shopping center	0
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	0
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	0
Tyumen TC "Green Coast "	0
Ufa TC" Central "	0
Ufa TC" Family "2	0
Volga TC "Volga Mall"	0
Vologda SEC "Marmelad"	0
Voronezh (Plekhanovskaya, 13)	0
Voronezh SEC "Maksimir"	0
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	0
Yakutsk TC" Central "	0
Yaroslavl TC" Altair "	0
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	1
! Yakutsk TC "Central" fran	0

item_category_name	Accessories - PS3 \
shop_name	
Adygea TC "Mega"	338
Balashikha TC "Oktyabr-Kinomir"	199
Chekhov TC" Karnaval "	319
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	213

Kaluga TC "XXI century"	337
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	218
Khimki TC" Mega "	814
Kolomna TC "Rio"	418
Krasnoyarsk TC "June"	689
Krasnoyarsk TC "Vzletka Plaza"	191
Kursk TC " Pushkinskiy "	424
Moscow MTRC" Afi Mall "	244
Moscow Shop C21	119
Moscow TC" MEGA Belya Dacha II "	868
Moscow TC" MEGA Teply Stan "II	986
Moscow TC" New Age "(Novokosino)	461
Moscow TC" Perlovsky "	360
Moscow TC" Semenovsky "	1154
Moscow TC" Serebryany House "	38
Moscow TEC" Atrium "	972
Moscow TK" Budenovskiy "(pav.A2)	71
Moscow TK" Budenovskiy "(pav.K7)	368
Moscow TTS" Areal "(Belyaev)	307
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	41
N.Novgorod SEC" Fantasy "	382
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	226
Omsk TC" Mega "	485
Outbound Trade	22
RostovNaDon Mega shopping center	309
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	46
Samara Melody shopping center	128
Samara ParkHouse shopping center	210
Sergiev Posad 7Ya shopping center	472
St. Petersburg Nevsky Center shopping center	548
St. Petersburg Sennaya shopping center	354
Surgut SEC "City Mall"	552
Tomsk SEC "Emerald City"	104
Tyumen SEC "Crystal"	102
Tyumen SEC "Goodwin"	665
Tyumen TC "Green Coast "	222
Ufa TC" Central "	244
Ufa TC" Family "2	329
Volga TC "Volga Mall"	301
Vologda SEC "Marmelad"	229
Voronezh (Plekhanovskaya, 13)	414
Voronezh SEC "Maksimir"	447
Voronezh SEC City-Park "Grad"	19

Yakutsk Ordzhonikidze, 56	592
Yakutsk TC" Central "	308
Yaroslavl TC" Altair "	218
Zhukovsky st. Chkalov 39m?	134
Zhukovsky st. Chkalov 39m ²	2
! Yakutsk Ordzhonikidze, 56 francs	93
! Yakutsk TC "Central" fran	58

item_category_name	Accessories - PS4 \
shop_name	
Adygea TC "Mega"	470
Balashikha TC "Oktyabr-Kinomir"	344
Chekhov TC" Karnaval "	425
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	514
Kaluga TC "XXI century"	460
Kazan TC "Behetle"	8
Kazan TC "ParkHouse" II	389
Khimki TC" Mega "	868
Kolomna TC "Rio"	469
Krasnoyarsk TC "June"	636
Krasnoyarsk TC "Vzletka Plaza"	218
Kursk TC " Pushkinskiy "	457
Moscow MTRC" Afi Mall "	581
Moscow Shop C21	322
Moscow TC" MEGA Belya Dacha II "	1076
Moscow TC" MEGA Teply Stan "II	1229
Moscow TC" New Age "(Novokosino)	514
Moscow TC" Perlovsky "	373
Moscow TC" Semenovsky "	1614
Moscow TC" Serebryany House "	13
Moscow TEC" Atrium "	1539
Moscow TK" Budenovskiy "(pav.A2)	11
Moscow TK" Budenovskiy "(pav.K7)	496
Moscow TTS" Areal "(Belyaev)	492
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	66
N.Novgorod SEC" Fantasy "	563
N.Novgorod SEC" RIO "	83
Novosibirsk SEC" Gallery Novosibirsk "	6
Novosibirsk SEC" Mega "	357
Omsk TC" Mega "	532
Outbound Trade	34
RostovNaDon Mega shopping center	295
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	1
RostovNaDonu SEC" Megacenter Horizon	141
Samara Melody shopping center	231
Samara ParkHouse shopping center	267

Sergiev Posad 7Ya shopping center	645
St. Petersburg Nevsky Center shopping center	1117
St. Petersburg Sennaya shopping center	487
Surgut SEC "City Mall"	862
Tomsk SEC "Emerald City"	355
Tyumen SEC "Crystal"	239
Tyumen SEC "Goodwin"	638
Tyumen TC "Green Coast "	237
Ufa TC" Central "	358
Ufa TC" Family "2	426
Volga TC "Volga Mall"	317
Vologda SEC "Marmelad"	313
Voronezh (Plekhanovskaya, 13)	644
Voronezh SEC "Maksimir"	508
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	786
Yakutsk TC" Central "	512
Yaroslavl TC" Altair "	370
Zhukovsky st. Chkalov 39m?	196
Zhukovsky st. Chkalov 39m ²	4
! Yakutsk Ordzhonikidze, 56 francs	35
! Yakutsk TC "Central" fran	15

item_category_name	Accessories - PSP \
shop_name	
Adygea TC "Mega"	39
Balashikha TC "Oktyabr-Kinomir"	28
Chekhov TC" Karnaval "	45
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	9
Kaluga TC "XXI century"	62
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	19
Khimki TC" Mega "	98
Kolomna TC "Rio"	136
Krasnoyarsk TC "June"	49
Krasnoyarsk TC "Vzletka Plaza"	17
Kursk TC " Pushkinskiy "	59
Moscow MTRC" Afi Mall "	17
Moscow Shop C21	11
Moscow TC" MEGA Belya Dacha II "	95
Moscow TC" MEGA Teply Stan "II	61
Moscow TC" New Age "(Novokosino)	52
Moscow TC" Perlovsky "	68
Moscow TC" Semenovsky "	136
Moscow TC" Serebryany House "	9
Moscow TEC" Atrium "	45
Moscow TK" Budenovskiy "(pav.A2)	4

Moscow TK" Budenovskiy "(pav.K7)	38
Moscow TTS" Areal "(Belyaev)	21
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	1
N.Novgorod SEC" Fantasy "	19
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	19
Omsk TC" Mega "	38
Outbound Trade	0
RostovNaDon Mega shopping center	44
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	2
Samara Melody shopping center	4
Samara ParkHouse shopping center	23
Sergiev Posad 7Ya shopping center	41
St. Petersburg Nevsky Center shopping center	35
St. Petersburg Sennaya shopping center	36
Surgut SEC "City Mall"	85
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	12
Tyumen SEC "Goodwin"	90
Tyumen TC "Green Coast "	55
Ufa TC" Central "	55
Ufa TC" Family "2	68
Volga TC "Volga Mall"	90
Vologda SEC "Marmelad"	39
Voronezh (Plekhanovskaya, 13)	22
Voronezh SEC "Maksimir"	90
Voronezh SEC City-Park "Grad"	2
Yakutsk Ordzhonikidze, 56	115
Yakutsk TC" Central "	73
Yaroslavl TC" Altair "	52
Zhukovsky st. Chkalov 39m?	6
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	35
! Yakutsk TC "Central" fran	23

item_category_name	Accessories - PSVita \
shop_name	
Adygea TC "Mega"	49
Balashikha TC "Oktyabr-Kinomir"	107
Chekhov TC" Karnaval "	153
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	100
Kaluga TC "XXI century"	66
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	48

Khimki TC" Mega "	275
Kolomna TC "Rio"	176
Krasnoyarsk TC "June"	208
Krasnoyarsk TC "Vzletka Plaza"	64
Kursk TC " Pushkinskiy "	136
Moscow MTRC" Afi Mall "	169
Moscow Shop C21	115
Moscow TC" MEGA Belya Dacha II "	403
Moscow TC" MEGA Teply Stan "II	306
Moscow TC" New Age "(Novokosino)	168
Moscow TC" Perlovsky "	167
Moscow TC" Semenovsky "	508
Moscow TC" Serebryany House "	21
Moscow TEC" Atrium "	449
Moscow TK" Budenovskiy "(pav.A2)	20
Moscow TK" Budenovskiy "(pav.K7)	160
Moscow TTS" Areal "(Belyaev)	163
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	15
N.Novgorod SEC" Fantasy "	148
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	87
Omsk TC" Mega "	65
Outbound Trade	0
RostovNaDon Mega shopping center	83
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	19
Samara Melody shopping center	66
Samara ParkHouse shopping center	102
Sergiev Posad 7Ya shopping center	136
St. Petersburg Nevsky Center shopping center	389
St. Petersburg Sennaya shopping center	120
Surgut SEC "City Mall"	167
Tomsk SEC "Emerald City"	35
Tyumen SEC "Crystal"	39
Tyumen SEC "Goodwin"	242
Tyumen TC "Green Coast "	59
Ufa TC" Central "	117
Ufa TC" Family "2	113
Volga TC "Volga Mall"	155
Vologda SEC "Marmelad"	95
Voronezh (Plekhanovskaya, 13)	169
Voronezh SEC "Maksimir"	200
Voronezh SEC City-Park "Grad"	7
Yakutsk Ordzhonikidze, 56	231
Yakutsk TC" Central "	129
Yaroslavl TC" Altair "	52

Zhukovsky st. Chkalov 39m?	48
Zhukovsky st. Chkalov 39m ²	1
! Yakutsk Ordzhonikidze, 56 francs	60
! Yakutsk TC "Central" fran	24
item_category_name	Accessories - XBOX 360 \
shop_name	
Adygea TC "Mega"	317
Balashikha TC "Oktyabr-Kinomir"	243
Chekhov TC" Karnaval "	492
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	132
Kaluga TC "XXI century"	443
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	231
Khimki TC" Mega "	691
Kolomna TC "Rio"	439
Krasnoyarsk TC "June"	368
Krasnoyarsk TC "Vzletka Plaza"	125
Kursk TC " Pushkinskiy "	330
Moscow MTRC" Afi Mall "	266
Moscow Shop C21	121
Moscow TC" MEGA Belya Dacha II "	988
Moscow TC" MEGA Teply Stan "II	963
Moscow TC" New Age "(Novokosino)	288
Moscow TC" Perlovsky "	362
Moscow TC" Semenovsky "	1096
Moscow TC" Serebryany House "	26
Moscow TEC" Atrium "	888
Moscow TK" Budenovskiy "(pav.A2)	37
Moscow TK" Budenovskiy "(pav.K7)	430
Moscow TTS" Areal "(Belyaev)	307
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	37
N.Novgorod SEC" Fantasy "	338
N.Novgorod SEC" RIO "	8
Novosibirsk SEC" Gallery Novosibirsk "	2
Novosibirsk SEC" Mega "	208
Omsk TC" Mega "	467
Outbound Trade	21
RostovNaDon Mega shopping center	334
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	57
Samara Melody shopping center	108
Samara ParkHouse shopping center	211
Sergiev Posad 7Ya shopping center	569
St. Petersburg Nevsky Center shopping center	520
St. Petersburg Sennaya shopping center	349

Surgut SEC "City Mall"	355
Tomsk SEC "Emerald City"	179
Tyumen SEC "Crystal"	145
Tyumen SEC "Goodwin"	571
Tyumen TC "Green Coast "	174
Ufa TC" Central "	337
Ufa TC" Family "2	481
Volga TC "Volga Mall"	319
Vologda SEC "Marmelad"	207
Voronezh (Plekhanovskaya, 13)	352
Voronezh SEC "Maksimir"	419
Voronezh SEC City-Park "Grad"	11
Yakutsk Ordzhonikidze, 56	826
Yakutsk TC" Central "	655
Yaroslavl TC" Altair "	247
Zhukovsky st. Chkalov 39m?	138
Zhukovsky st. Chkalov 39m ²	6
! Yakutsk Ordzhonikidze, 56 francs	116
! Yakutsk TC "Central" fran	57

item_category_name	Accessories - XBOX ONE \
shop_name	
Adygea TC "Mega"	68
Balashikha TC "Oktyabr-Kinomir"	60
Chekhov TC" Karnaval "	57
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	161
Kaluga TC "XXI century"	87
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	58
Khimki TC" Mega "	123
Kolomna TC "Rio"	54
Krasnoyarsk TC "June"	77
Krasnoyarsk TC "Vzletka Plaza"	28
Kursk TC " Pushkinskiy "	45
Moscow MTRC" Afi Mall "	153
Moscow Shop C21	74
Moscow TC" MEGA Belya Dacha II "	206
Moscow TC" MEGA Teply Stan "II	238
Moscow TC" New Age "(Novokosino)	53
Moscow TC" Perlovsky "	34
Moscow TC" Semenovsky "	212
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	289
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	65
Moscow TTS" Areal "(Belyaev)	119
Moscow" Sale "	0

Mytishchi SEC" XL-3 "	23
N.Novgorod SEC" Fantasy "	131
N.Novgorod SEC" RIO "	40
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	87
Omsk TC" Mega "	122
Outbound Trade	15
RostovNaDon Mega shopping center	46
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	80
Samara Melody shopping center	47
Samara ParkHouse shopping center	56
Sergiev Posad 7Ya shopping center	79
St. Petersburg Nevsky Center shopping center	271
St. Petersburg Sennaya shopping center	72
Surgut SEC "City Mall"	99
Tomsk SEC "Emerald City"	94
Tyumen SEC "Crystal"	56
Tyumen SEC "Goodwin"	71
Tyumen TC "Green Coast "	28
Ufa TC" Central "	56
Ufa TC" Family "2	73
Volga TC "Volga Mall"	45
Vologda SEC "Marmelad"	65
Voronezh (Plekhanovskaya, 13)	92
Voronezh SEC "Maksimir"	63
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	168
Yakutsk TC" Central "	81
Yaroslavl TC" Altair "	72
Zhukovsky st. Chkalov 39m?	36
Zhukovsky st. Chkalov 39m ²	1
! Yakutsk Ordzhonikidze, 56 francs	0
! Yakutsk TC "Central" fran	0

item_category_name	Android games - Number \
shop_name	
Adygea TC "Mega"	0
Balashikha TC "Oktyabr-Kinomir"	0
Chekhov TC" Karnaval "	0
Digital warehouse 1C-Online	91
Internet-shop of emergency situations	0
Kaluga TC "XXI century"	0
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	0
Khimki TC" Mega "	0
Kolomna TC "Rio"	0
Krasnoyarsk TC "June"	0

Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	0
Moscow MTRC" Afi Mall "	0
Moscow Shop C21	0
Moscow TC" MEGA Belya Dacha II "	0
Moscow TC" MEGA Teply Stan "II	0
Moscow TC" New Age "(Novokosino)	0
Moscow TC" Perlovsky "	0
Moscow TC" Semenovsky "	0
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	0
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	0
Moscow TTS" Areal "(Belyaev)	0
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	0
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	0
Omsk TC" Mega "	0
Outbound Trade	0
RostovNaDon Mega shopping center	0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	0
Samara Melody shopping center	0
Samara ParkHouse shopping center	0
Sergiev Posad 7Ya shopping center	0
St. Petersburg Nevsky Center shopping center	0
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	0
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	0
Tyumen TC "Green Coast "	0
Ufa TC" Central "	0
Ufa TC" Family "2	0
Volga TC "Volga Mall"	0
Vologda SEC "Marmelad"	0
Voronezh (Plekhanovskaya, 13)	0
Voronezh SEC "Maksimir"	0
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	0
Yakutsk TC" Central "	0
Yaroslavl TC" Altair "	0
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	0

! Yakutsk TC "Central" fran

0

item_category_name shop_name	Batteries \
Adygea TC "Mega"	34
Balashikha TC "Oktyabr-Kinomir"	36
Chekhov TC" Karnaval "	152
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	2
Kaluga TC "XXI century"	62
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	164
Khimki TC" Mega "	169
Kolomna TC "Rio"	202
Krasnoyarsk TC "June"	122
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	96
Moscow MTRC" Afi Mall "	129
Moscow Shop C21	84
Moscow TC" MEGA Belya Dacha II "	53
Moscow TC" MEGA Teply Stan "II	411
Moscow TC" New Age "(Novokosino)	108
Moscow TC" Perlovsky "	130
Moscow TC" Semenovsky "	439
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	338
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	44
Moscow TTS" Areal "(Belyaev)	73
Moscow" Sale "	2
Mytishchi SEC" XL-3 "	29
N.Novgorod SEC" Fantasy "	67
N.Novgorod SEC" RIO "	43
Novosibirsk SEC" Gallery Novosibirsk "	2
Novosibirsk SEC" Mega "	157
Omsk TC" Mega "	74
Outbound Trade	0
RostovNaDon Mega shopping center	75
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	73
Samara Melody shopping center	546
Samara ParkHouse shopping center	136
Sergiev Posad 7Ya shopping center	128
St. Petersburg Nevsky Center shopping center	155
St. Petersburg Sennaya shopping center	155
Surgut SEC "City Mall"	134
Tomsk SEC "Emerald City"	49
Tyumen SEC "Crystal"	36

Tyumen SEC "Goodwin"	241
Tyumen TC "Green Coast "	242
Ufa TC" Central "	183
Ufa TC" Family "2	117
Volga TC "Volga Mall"	50
Vologda SEC "Marmelad"	82
Voronezh (Plekhanovskaya, 13)	392
Voronezh SEC "Maksimir"	110
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	374
Yakutsk TC" Central "	170
Yaroslavl TC" Altair "	347
Zhukovsky st. Chkalov 39m?	89
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	48
! Yakutsk TC "Central" fran	41
 item_category_name	Blank media (piece) \
shop_name	
Adygea TC "Mega"	42
Balashikha TC "Oktyabr-Kinomir"	0
Chekhov TC" Karnaval "	310
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	3
Kaluga TC "XXI century"	0
Kazan TC "Behetle"	105
Kazan TC "ParkHouse" II	0
Khimki TC" Mega "	120
Kolomna TC "Rio"	186
Krasnoyarsk TC "June"	86
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	0
Moscow MTRC" Afi Mall "	0
Moscow Shop C21	101
Moscow TC" MEGA Belaya Dacha II "	0
Moscow TC" MEGA Teply Stan "II	0
Moscow TC" New Age "(Novokosino)	285
Moscow TC" Perlovsky "	266
Moscow TC" Semenovsky "	0
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	254
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	0
Moscow TTS" Areal "(Belyaev)	0
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	0
N.Novgorod SEC" RIO "	0

Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	84
Omsk TC" Mega "	0
Outbound Trade	0
RostovNaDon Mega shopping center	67
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	0
Samara Melody shopping center	548
Samara ParkHouse shopping center	0
Sergiev Posad 7Ya shopping center	0
St. Petersburg Nevsky Center shopping center	0
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	0
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	0
Tyumen TC "Green Coast "	508
Ufa TC" Central "	291
Ufa TC" Family "2	0
Volga TC "Volga Mall"	208
Vologda SEC "Marmelad"	0
Voronezh (Plekhanovskaya, 13)	290
Voronezh SEC "Maksimir"	0
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	228
Yakutsk TC" Central "	170
Yaroslavl TC" Altair "	0
Zhukovsky st. Chkalov 39m?	110
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	69
! Yakutsk TC "Central" fran	59

item_category_name	... \
shop_name	...
Adygea TC "Mega"	...
Balashikha TC "Oktyabr-Kinomir"	...
Chekhov TC" Karnaval "	...
Digital warehouse 1C-Online	...
Internet-shop of emergency situations	...
Kaluga TC "XXI century"	...
Kazan TC "Behetle"	...
Kazan TC "ParkHouse" II	...
Khimki TC" Mega "	...
Kolomna TC "Rio"	...
Krasnoyarsk TC "June"	...
Krasnoyarsk TC "Vzletka Plaza"	...
Kursk TC " Pushkinskiy "	...
Moscow MTRC" Afi Mall "	...

Moscow Shop C21	...
Moscow TC" MEGA Belya Dacha II "	...
Moscow TC" MEGA Teply Stan "II	...
Moscow TC" New Age "(Novokosino)	...
Moscow TC" Perlovsky "	...
Moscow TC" Semenovsky "	...
Moscow TC" Serebryany House "	...
Moscow TEC" Atrium "	...
Moscow TK" Budenovskiy "(pav.A2)	...
Moscow TK" Budenovskiy "(pav.K7)	...
Moscow TTS" Areal "(Belyaev)	...
Moscow" Sale "	...
Mytishchi SEC" XL-3 "	...
N.Novgorod SEC" Fantasy "	...
N.Novgorod SEC" RIO "	...
Novosibirsk SEC" Gallery Novosibirsk "	...
Novosibirsk SEC" Mega "	...
Omsk TC" Mega "	...
Outbound Trade	...
RostovNaDon Mega shopping center	...
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	...
RostovNaDonu SEC" Megacenter Horizon	...
Samara Melody shopping center	...
Samara ParkHouse shopping center	...
Sergiev Posad 7Ya shopping center	...
St. Petersburg Nevsky Center shopping center	...
St. Petersburg Sennaya shopping center	...
Surgut SEC "City Mall"	...
Tomsk SEC "Emerald City"	...
Tyumen SEC "Crystal"	...
Tyumen SEC "Goodwin"	...
Tyumen TC "Green Coast "	...
Ufa TC" Central "	...
Ufa TC" Family "2	...
Volga TC "Volga Mall"	...
Vologda SEC "Marmelad"	...
Voronezh (Plekhanovskaya, 13)	...
Voronezh SEC "Maksimir"	...
Voronezh SEC City-Park "Grad"	...
Yakutsk Ordzhonikidze, 56	...
Yakutsk TC" Central "	...
Yaroslavl TC" Altair "	...
Zhukovsky st. Chkalov 39m?	...
Zhukovsky st. Chkalov 39m ²	...
! Yakutsk Ordzhonikidze, 56 francs	...
! Yakutsk TC "Central" fran	...

\		
shop_name		
Adygea TC "Mega"	165	
Balashikha TC "Oktyabr-Kinomir"	62	
Chekhov TC" Karnaval "	136	
Digital warehouse 1C-Online	0	
Internet-shop of emergency situations	1447	
Kaluga TC "XXI century"	115	
Kazan TC "Behetle"	8	
Kazan TC "ParkHouse" II	128	
Khimki TC" Mega "	402	
Kolomna TC "Rio"	222	
Krasnoyarsk TC "June"	102	
Krasnoyarsk TC "Vzletka Plaza"	30	
Kursk TC " Pushkinskiy "	113	
Moscow MTRC" Afi Mall "	206	
Moscow Shop C21	2134	
Moscow TC" MEGA Belya Dacha II "	833	
Moscow TC" MEGA Teply Stan "II	387	
Moscow TC" New Age "(Novokosino)	260	
Moscow TC" Perlovsky "	239	
Moscow TC" Semenovsky "	641	
Moscow TC" Serebryany House "	52	
Moscow TEC" Atrium "	784	
Moscow TK" Budenovskiy "(pav.A2)	49	
Moscow TK" Budenovskiy "(pav.K7)	520	
Moscow TTS" Areal "(Belyaev)	389	
Moscow" Sale "	0	
Mytishchi SEC" XL-3 "	9	
N.Novgorod SEC" Fantasy "	86	
N.Novgorod SEC" RIO "	19	
Novosibirsk SEC" Gallery Novosibirsk "	0	
Novosibirsk SEC" Mega "	171	
Omsk TC" Mega "	135	
Outbound Trade	0	
RostovNaDon Mega shopping center	95	
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0	
RostovNaDonu SEC" Megacenter Horizon	27	
Samara Melody shopping center	68	
Samara ParkHouse shopping center	101	
Sergiev Posad 7Ya shopping center	144	
St. Petersburg Nevsky Center shopping center	430	
St. Petersburg Sennaya shopping center	241	
Surgut SEC "City Mall"	183	
Tomsk SEC "Emerald City"	26	
Tyumen SEC "Crystal"	29	
Tyumen SEC "Goodwin"	79	
Tyumen TC "Green Coast "	49	

Ufa TC" Central "	114
Ufa TC" Family "2	171
Volga TC "Volga Mall"	87
Vologda SEC "Marmelad"	47
Voronezh (Plekhanovskaya, 13)	235
Voronezh SEC "Maksimir"	129
Voronezh SEC City-Park "Grad"	6
Yakutsk Ordzhonikidze, 56	214
Yakutsk TC" Central "	120
Yaroslavl TC" Altair "	106
Zhukovsky st. Chkalov 39m?	130
Zhukovsky st. Chkalov 39m ²	3
! Yakutsk Ordzhonikidze, 56 francs	20
! Yakutsk TC "Central" fran	14

item_category_name	Programs - Educational \
shop_name	
Adygea TC "Mega"	33
Balashikha TC "Oktyabr-Kinomir"	12
Chekhov TC" Karnaval "	126
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	140
Kaluga TC "XXI century"	24
Kazan TC "Behetle"	5
Kazan TC "ParkHouse" II	7
Khimki TC" Mega "	118
Kolomna TC "Rio"	126
Krasnoyarsk TC "June"	60
Krasnoyarsk TC "Vzletka Plaza"	4
Kursk TC " Pushkinskiy "	22
Moscow MTRC" Afi Mall "	52
Moscow Shop C21	543
Moscow TC" MEGA Belya Dacha II "	5
Moscow TC" MEGA Teply Stan "II	69
Moscow TC" New Age "(Novokosino)	61
Moscow TC" Perlovsky "	49
Moscow TC" Semenovsky "	29
Moscow TC" Serebryany House "	48
Moscow TEC" Atrium "	227
Moscow TK" Budenovskiy "(pav.A2)	54
Moscow TK" Budenovskiy "(pav.K7)	66
Moscow TTS" Areal "(Belyaev)	12
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	79
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	108

Omsk TC" Mega "	130
Outbound Trade	0
RostovNaDon Mega shopping center	59
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	1
Samara Melody shopping center	89
Samara ParkHouse shopping center	11
Sergiev Posad 7Ya shopping center	25
St. Petersburg Nevsky Center shopping center	30
St. Petersburg Sennaya shopping center	8
Surgut SEC "City Mall"	1
Tomsk SEC "Emerald City"	3
Tyumen SEC "Crystal"	1
Tyumen SEC "Goodwin"	7
Tyumen TC "Green Coast "	85
Ufa TC" Central "	153
Ufa TC" Family "2	8
Volga TC "Volga Mall"	103
Vologda SEC "Marmelad"	12
Voronezh (Plekhanovskaya, 13)	176
Voronezh SEC "Maksimir"	17
Voronezh SEC City-Park "Grad"	20
Yakutsk Ordzhonikidze, 56	324
Yakutsk TC" Central "	172
Yaroslavl TC" Altair "	4
Zhukovsky st. Chkalov 39m?	8
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	104
! Yakutsk TC "Central" fran	61

item_category_name	Programs - Educational
(Digit) \	
shop_name	
Adygea TC "Mega"	
0	
Balashikha TC "Oktyabr-Kinomir"	
0	
Chekhov TC" Karnaval "	
0	
Digital warehouse 1C-Online	
2346	
Internet-shop of emergency situations	
0	
Kaluga TC "XXI century"	
0	
Kazan TC "Behetle"	
0	
Kazan TC "ParkHouse" II	

0
Khimki TC" Mega "
0
Kolomna TC "Rio"
0
Krasnoyarsk TC "June"
0
Krasnoyarsk TC "Vzletka Plaza"
0
Kursk TC " Pushkinskiy "
0
Moscow MTRC" Afi Mall "
0
Moscow Shop C21
0
Moscow TC" MEGA Belya Dacha II "
0
Moscow TC" MEGA Teply Stan "II
0
Moscow TC" New Age "(Novokosino)
0
Moscow TC" Perlovsky "
0
Moscow TC" Semenovsky "
0
Moscow TC" Serebryany House "
0
Moscow TEC" Atrium "
0
Moscow TK" Budenovskiy "(pav.A2)
0
Moscow TK" Budenovskiy "(pav.K7)
0
Moscow TTS" Areal "(Belyaev)
0
Moscow" Sale "
0
Mytishchi SEC" XL-3 "
0
N.Novgorod SEC" Fantasy "
0
N.Novgorod SEC" RIO "
0
Novosibirsk SEC" Gallery Novosibirsk "
0
Novosibirsk SEC" Mega "
0
Omsk TC" Mega "

0
Outbound Trade
0
RostovNaDon Mega shopping center
0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...
0
RostovNaDon SEC" Megacenter Horizon
0
Samara Melody shopping center
0
Samara ParkHouse shopping center
0
Sergiev Posad 7Ya shopping center
0
St. Petersburg Nevsky Center shopping center
0
St. Petersburg Sennaya shopping center
0
Surgut SEC "City Mall"
0
Tomsk SEC "Emerald City"
0
Tyumen SEC "Crystal"
0
Tyumen SEC "Goodwin"
0
Tyumen TC "Green Coast "
0
Ufa TC" Central "
0
Ufa TC" Family "2
0
Volga TC "Volga Mall"
0
Vologda SEC "Marmelad"
0
Voronezh (Plekhanovskaya, 13)
0
Voronezh SEC "Maksimir"
0
Voronezh SEC City-Park "Grad"
0
Yakutsk Ordzhonikidze, 56
0
Yakutsk TC" Central "
0
Yaroslavl TC" Altair "

0
Zhukovsky st. Chkalov 39^m?
0
Zhukovsky st. Chkalov 39^{m²}
0
! Yakutsk Ordzhonikidze, 56 francs
0
! Yakutsk TC "Central" fran
0

item_category_name Programs - For home and
office \
shop_name
Adygea TC "Mega"
481
Balashikha TC "Oktyabr-Kinomir"
152
Chekhov TC" Karnaval "
434
Digital warehouse 1C-Online
0
Internet-shop of emergency situations
417
Kaluga TC "XXI century"
1825
Kazan TC "Behetle"
38
Kazan TC "ParkHouse" II
328
Khimki TC" Mega "
1237
Kolomna TC "Rio"
1087
Krasnoyarsk TC "June"
515
Krasnoyarsk TC "Vzletka Plaza"
192
Kursk TC " Pushkinskiy "
761
Moscow MTRC" Afi Mall "
500
Moscow Shop C21
2773
Moscow TC" MEGA Belya Dacha II "
1486
Moscow TC" MEGA Teply Stan "II
804
Moscow TC" New Age "(Novokosino)

1694
Moscow TC" Perlovsky "
671
Moscow TC" Semenovsky "
3177
Moscow TC" Serebryany House "
168
Moscow TEC" Atrium "
1556
Moscow TK" Budenovskiy "(pav.A2)
501
Moscow TK" Budenovskiy "(pav.K7)
3780
Moscow TTS" Areal "(Belyaev)
909
Moscow" Sale "
0
Mytishchi SEC" XL-3 "
25
N.Novgorod SEC" Fantasy "
219
N.Novgorod SEC" RIO "
18
Novosibirsk SEC" Gallery Novosibirsk "
3
Novosibirsk SEC" Mega "
617
Omsk TC" Mega "
509
Outbound Trade
4
RostovNaDon Mega shopping center
603
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...
0
RostovNaDonu SEC" Megacenter Horizon
41
Samara Melody shopping center
608
Samara ParkHouse shopping center
388
Sergiev Posad 7Ya shopping center
845
St. Petersburg Nevsky Center shopping center
630
St. Petersburg Sennaya shopping center
330
Surgut SEC "City Mall"

568
Tomsk SEC "Emerald City"
56
Tyumen SEC "Crystal"
115
Tyumen SEC "Goodwin"
740
Tyumen TC "Green Coast "
542
Ufa TC" Central "
464
Ufa TC" Family "2
574
Volga TC "Volga Mall"
718
Vologda SEC "Marmelad"
122
Voronezh (Plekhanovskaya, 13)
1421
Voronezh SEC "Maksimir"
856
Voronezh SEC City-Park "Grad"
35
Yakutsk Ordzhonikidze, 56
2760
Yakutsk TC" Central "
1111
Yaroslavl TC" Altair "
742
Zhukovsky st. Chkalov 39m?
920
Zhukovsky st. Chkalov 39m²
20
! Yakutsk Ordzhonikidze, 56 francs
252
! Yakutsk TC "Central" fran
121

item_category_name
office (Digit) \
shop_name
Adygea TC "Mega"
0
Balashikha TC "Oktyabr-Kinomir"
0
Chekhov TC" Karnaval "
0
Digital warehouse 1C-Online

Programs - For home and

3746
Internet-shop of emergency situations
0
Kaluga TC "XXI century"
0
Kazan TC "Behetle"
0
Kazan TC "ParkHouse" II
0
Khimki TC" Mega "
0
Kolomna TC "Rio"
0
Krasnoyarsk TC "June"
0
Krasnoyarsk TC "Vzletka Plaza"
0
Kursk TC " Pushkinskiy "
0
Moscow MTRC" Afi Mall "
0
Moscow Shop C21
0
Moscow TC" MEGA Belya Dacha II "
0
Moscow TC" MEGA Teply Stan "II
0
Moscow TC" New Age "(Novokosino)
0
Moscow TC" Perlovsky "
0
Moscow TC" Semenovsky "
0
Moscow TC" Serebryany House "
0
Moscow TEC" Atrium "
0
Moscow TK" Budenovskiy "(pav.A2)
0
Moscow TK" Budenovskiy "(pav.K7)
0
Moscow TTS" Areal "(Belyaev)
0
Moscow" Sale "
0
Mytishchi SEC" XL-3 "
0
N.Novgorod SEC" Fantasy "

0
N.Novgorod SEC" RIO "
0
Novosibirsk SEC" Gallery Novosibirsk "
0
Novosibirsk SEC" Mega "
0
Omsk TC" Mega "
0
Outbound Trade
0
RostovNaDon Mega shopping center
0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...
0
RostovNaDonu SEC" Megacenter Horizon
0
Samara Melody shopping center
0
Samara ParkHouse shopping center
0
Sergiev Posad 7Ya shopping center
0
St. Petersburg Nevsky Center shopping center
0
St. Petersburg Sennaya shopping center
0
Surgut SEC "City Mall"
0
Tomsk SEC "Emerald City"
0
Tyumen SEC "Crystal"
0
Tyumen SEC "Goodwin"
0
Tyumen TC "Green Coast "
0
Ufa TC" Central "
0
Ufa TC" Family "2
0
Volga TC "Volga Mall"
0
Vologda SEC "Marmelad"
0
Voronezh (Plekhanovskaya, 13)
0
Voronezh SEC "Maksimir"

```

0
Voronezh SEC City-Park "Grad"
0
Yakutsk Ordzhonikidze, 56
0
Yakutsk TC" Central "
0
Yaroslavl TC" Altair "
0
Zhukovsky st. Chkalov 39m?
0
Zhukovsky st. Chkalov 39m2
0
! Yakutsk Ordzhonikidze, 56 francs
0
! Yakutsk TC "Central" fran
0

```

item_category_name	Programs - MAC (Digit) \
shop_name	
Adygea TC "Mega"	0
Balashikha TC "Oktyabr-Kinomir"	0
Chekhov TC" Karnaval "	0
Digital warehouse 1C-Online	56
Internet-shop of emergency situations	0
Kaluga TC "XXI century"	0
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	0
Khimki TC" Mega "	0
Kolomna TC "Rio"	0
Krasnoyarsk TC "June"	0
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	0
Moscow MTRC" Afi Mall "	0
Moscow Shop C21	0
Moscow TC" MEGA Belya Dacha II "	0
Moscow TC" MEGA Teply Stan "II	0
Moscow TC" New Age "(Novokosino)	0
Moscow TC" Perlovsky "	0
Moscow TC" Semenovsky "	0
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	0
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	0
Moscow TTS" Areal "(Belyaev)	0
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	0

N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	0
Omsk TC" Mega "	0
Outbound Trade	0
RostovNaDon Mega shopping center	0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	0
Samara Melody shopping center	0
Samara ParkHouse shopping center	0
Sergiev Posad 7Ya shopping center	0
St. Petersburg Nevsky Center shopping center	0
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	0
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	0
Tyumen TC "Green Coast "	0
Ufa TC" Central "	0
Ufa TC" Family "2	0
Volga TC "Volga Mall"	0
Vologda SEC "Marmelad"	0
Voronezh (Plekhanovskaya, 13)	0
Voronezh SEC "Maksimir"	0
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	0
Yakutsk TC" Central "	0
Yaroslavl TC" Altair "	0
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	0
! Yakutsk TC "Central" fran	0

item_category_name	Service \
shop_name	
Adygea TC "Mega"	161
Balashikha TC "Oktyabr-Kinomir"	179
Chekhov TC" Karnaval "	159
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	0
Kaluga TC "XXI century"	185
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	178
Khimki TC" Mega "	179
Kolomna TC "Rio"	218
Krasnoyarsk TC "June"	193
Krasnoyarsk TC "Vzletka Plaza"	62
Kursk TC " Pushkinskiy "	176

Moscow MTRC" Afi Mall "	218
Moscow Shop C21	262
Moscow TC" MEGA Belya Dacha II "	373
Moscow TC" MEGA Teply Stan "II	248
Moscow TC" New Age "(Novokosino)	162
Moscow TC" Perlovsky "	156
Moscow TC" Semenovsky "	478
Moscow TC" Serebryany House "	8
Moscow TEC" Atrium "	288
Moscow TK" Budenovskiy "(pav.A2)	8
Moscow TK" Budenovskiy "(pav.K7)	245
Moscow TTS" Areal "(Belyaev)	265
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	49
N.Novgorod SEC" Fantasy "	279
N.Novgorod SEC" RIO "	72
Novosibirsk SEC" Gallery Novosibirsk "	1
Novosibirsk SEC" Mega "	129
Omsk TC" Mega "	274
Outbound Trade	4
RostovNaDon Mega shopping center	97
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	4
RostovNaDonu SEC" Megacenter Horizon	104
Samara Melody shopping center	90
Samara ParkHouse shopping center	145
Sergiev Posad 7Ya shopping center	226
St. Petersburg Nevsky Center shopping center	382
St. Petersburg Sennaya shopping center	165
Surgut SEC "City Mall"	212
Tomsk SEC "Emerald City"	195
Tyumen SEC "Crystal"	105
Tyumen SEC "Goodwin"	117
Tyumen TC "Green Coast "	0
Ufa TC" Central "	156
Ufa TC" Family "2	223
Volga TC "Volga Mall"	302
Vologda SEC "Marmelad"	195
Voronezh (Plekhanovskaya, 13)	330
Voronezh SEC "Maksimir"	328
Voronezh SEC City-Park "Grad"	1
Yakutsk Ordzhonikidze, 56	160
Yakutsk TC" Central "	116
Yaroslavl TC" Altair "	205
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	0
! Yakutsk TC "Central" fran	0

item_category_name	Service - Tickets \
shop_name	
Adygea TC "Mega"	0
Balashikha TC "Oktyabr-Kinomir"	44
Chekhov TC" Karnaval "	36
Digital warehouse 1C-Online	3
Internet-shop of emergency situations	169
Kaluga TC "XXI century"	46
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	18
Khimki TC" Mega "	0
Kolomna TC "Rio"	73
Krasnoyarsk TC "June"	4
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	20
Moscow MTRC" Afi Mall "	65
Moscow Shop C21	53
Moscow TC" MEGA Belya Dacha II "	19
Moscow TC" MEGA Teply Stan "II	109
Moscow TC" New Age "(Novokosino)	1
Moscow TC" Perlovsky "	0
Moscow TC" Semenovsky "	134
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	144
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	41
Moscow TTS" Areal "(Belyaev)	61
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	30
N.Novgorod SEC" RIO "	18
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	2
Omsk TC" Mega "	5
Outbound Trade	0
RostovNaDon Mega shopping center	3
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	9
Samara Melody shopping center	10
Samara ParkHouse shopping center	10
Sergiev Posad 7Ya shopping center	65
St. Petersburg Nevsky Center shopping center	43
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	3
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	10
Tyumen TC "Green Coast "	0

Ufa TC" Central "	2
Ufa TC" Family "2	4
Volga TC "Volga Mall"	2
Vologda SEC "Marmelad"	11
Voronezh (Plekhanovskaya, 13)	32
Voronezh SEC "Maksimir"	13
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	0
Yakutsk TC" Central "	0
Yaroslavl TC" Altair "	12
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	0
! Yakutsk TC "Central" fran	0

item_category_name	Tickets (Digital) \
shop_name	
Adygea TC "Mega"	3
Balashikha TC "Oktyabr-Kinomir"	35
Chekhov TC" Karnaval "	41
Digital warehouse 1C-Online	507
Internet-shop of emergency situations	111
Kaluga TC "XXI century"	46
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	9
Khimki TC" Mega "	76
Kolomna TC "Rio"	66
Krasnoyarsk TC "June"	0
Krasnoyarsk TC "Vzletka Plaza"	1
Kursk TC " Pushkinskiy "	21
Moscow MTRC" Afi Mall "	57
Moscow Shop C21	55
Moscow TC" MEGA Belya Dacha II "	82
Moscow TC" MEGA Teply Stan "II	57
Moscow TC" New Age "(Novokosino)	44
Moscow TC" Perlovsky "	70
Moscow TC" Semenovsky "	114
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	98
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	39
Moscow TTS" Areal "(Belyaev)	62
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	8
N.Novgorod SEC" Fantasy "	25
N.Novgorod SEC" RIO "	5
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	1

Omsk TC" Mega "	2
Outbound Trade	0
RostovNaDon Mega shopping center	9
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDon SEC" Megacenter Horizon	7
Samara Melody shopping center	4
Samara ParkHouse shopping center	7
Sergiev Posad 7Ya shopping center	52
St. Petersburg Nevsky Center shopping center	26
St. Petersburg Sennaya shopping center	22
Surgut SEC "City Mall"	2
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	2
Tyumen TC "Green Coast "	0
Ufa TC" Central "	3
Ufa TC" Family "2	8
Volga TC "Volga Mall"	2
Vologda SEC "Marmelad"	26
Voronezh (Plekhanovskaya, 13)	25
Voronezh SEC "Maksimir"	8
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	5
Yakutsk TC" Central "	1
Yaroslavl TC" Altair "	27
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	0
! Yakutsk TC "Central" fran	0

item_category_name	PC - Headsets / Headphones
shop_name	
Adygea TC "Mega"	0
Balashikha TC "Oktyabr-Kinomir"	0
Chekhov TC" Karnaval "	0
Digital warehouse 1C-Online	0
Internet-shop of emergency situations	0
Kaluga TC "XXI century"	0
Kazan TC "Behetle"	0
Kazan TC "ParkHouse" II	0
Khimki TC" Mega "	0
Kolomna TC "Rio"	0
Krasnoyarsk TC "June"	0
Krasnoyarsk TC "Vzletka Plaza"	0
Kursk TC " Pushkinskiy "	0
Moscow MTRC" Afi Mall "	0
Moscow Shop C21	0
Moscow TC" MEGA Belya Dacha II "	0

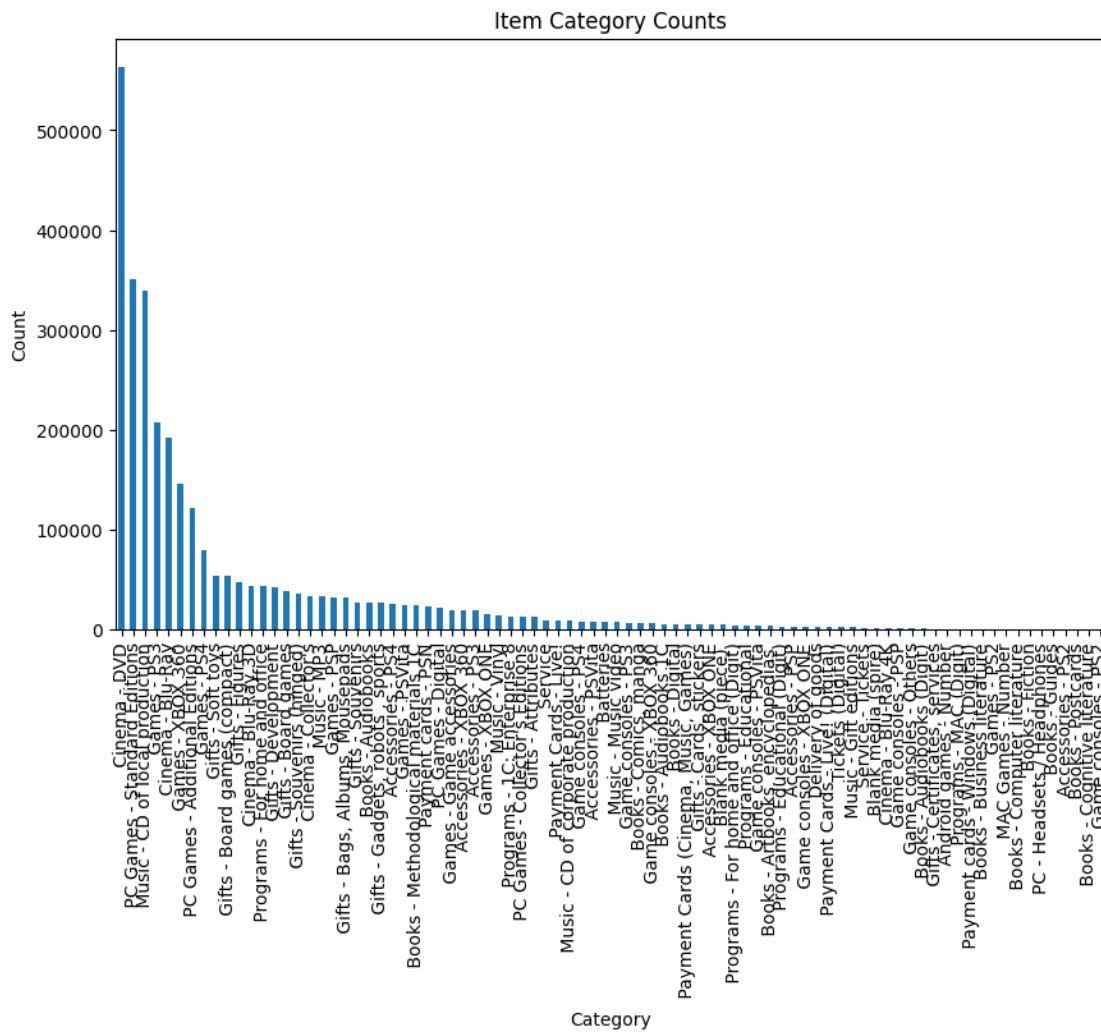
Moscow TC" MEGA Teply Stan "II	0
Moscow TC" New Age "(Novokosino)	0
Moscow TC" Perlovsky "	0
Moscow TC" Semenovsky "	0
Moscow TC" Serebryany House "	0
Moscow TEC" Atrium "	1
Moscow TK" Budenovskiy "(pav.A2)	0
Moscow TK" Budenovskiy "(pav.K7)	0
Moscow TTS" Areal "(Belyaev)	0
Moscow" Sale "	0
Mytishchi SEC" XL-3 "	0
N.Novgorod SEC" Fantasy "	0
N.Novgorod SEC" RIO "	0
Novosibirsk SEC" Gallery Novosibirsk "	0
Novosibirsk SEC" Mega "	0
Omsk TC" Mega "	1
Outbound Trade	0
RostovNaDon Mega shopping center	0
RostovNaDon Megatsentr Gorizont Ostrovnoy shop...	0
RostovNaDonu SEC" Megacenter Horizon	0
Samara Melody shopping center	0
Samara ParkHouse shopping center	0
Sergiev Posad 7Ya shopping center	0
St. Petersburg Nevsky Center shopping center	0
St. Petersburg Sennaya shopping center	0
Surgut SEC "City Mall"	0
Tomsk SEC "Emerald City"	0
Tyumen SEC "Crystal"	0
Tyumen SEC "Goodwin"	0
Tyumen TC "Green Coast "	0
Ufa TC" Central "	0
Ufa TC" Family "2	0
Volga TC "Volga Mall"	0
Vologda SEC "Marmelad"	0
Voronezh (Plekhanovskaya, 13)	0
Voronezh SEC "Maksimir"	0
Voronezh SEC City-Park "Grad"	0
Yakutsk Ordzhonikidze, 56	0
Yakutsk TC" Central "	0
Yaroslavl TC" Altair "	0
Zhukovsky st. Chkalov 39m?	0
Zhukovsky st. Chkalov 39m ²	0
! Yakutsk Ordzhonikidze, 56 francs	1
! Yakutsk TC "Central" fran	0

[60 rows x 84 columns]

Frequency of 'Low' price range: 0
item_category_name

Accessories - PS2	215.500000
Accessories - PS3	2255.981900
Accessories - PS4	2105.419083
Accessories - PSP	555.057055
Accessories - PSVita	864.652948
	...
Programs - MAC (Digit)	3403.571429
Service	1094.625747
Service - Tickets	1246.023391
Tickets (Digital)	1344.607752
PC - Headsets / Headphones	87.800000

Name: item_price, Length: 84, dtype: float64



[]: #stationarity analysis

```
[ ]: # Convert the date column to datetime format
final_dataset['date'] = pd.to_datetime(final_dataset['date'], format='%d.%m.%Y')

monthly_data = final_dataset.groupby(final_dataset['date'].dt.to_period('M')).
    agg({
        'item_cnt_month': 'sum',
    }).reset_index()

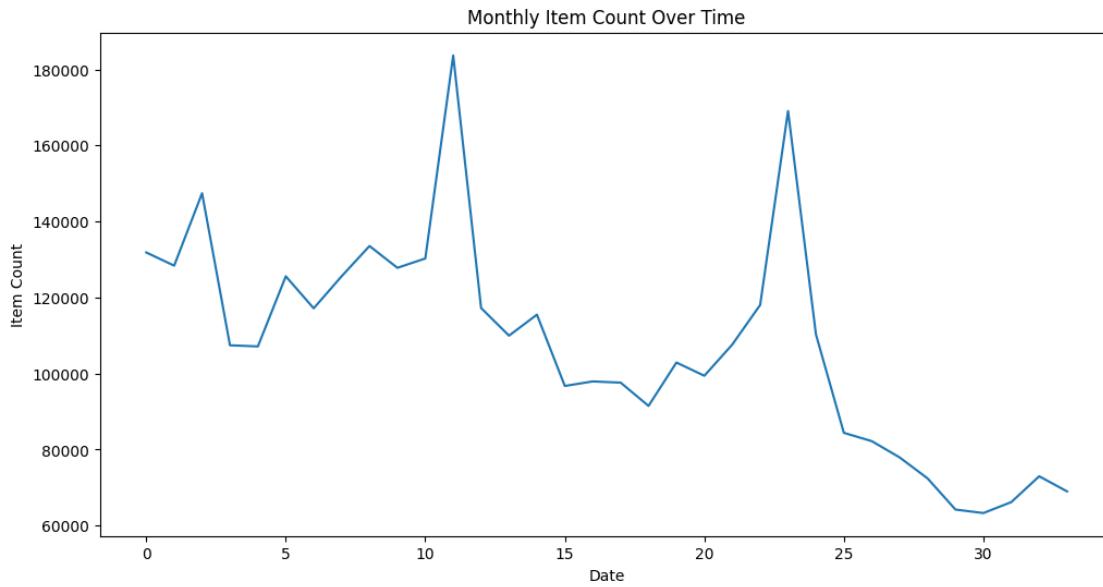
def adf_test(timeseries):
    result = adfuller(timeseries, autolag='AIC')
    print('ADF Statistic:', result[0])
    print('p-value:', result[1])
    print('Critical Values:')
    for key, value in result[4].items():
        print(f' {key}: {value}')

    if result[1] <= 0.05:
        print("Stationary (Reject the null hypothesis)")
    else:
        print("Non-Stationary (Fail to reject the null hypothesis)")

item_cnt_month_series = monthly_data['item_cnt_month']

plt.figure(figsize=(12, 6))
plt.plot(item_cnt_month_series)
plt.title('Monthly Item Count Over Time')
plt.xlabel('Date')
plt.ylabel('Item Count')
plt.show()

adf_test(item_cnt_month_series)
```



ADF Statistic: -2.372251519825604

p-value: 0.1497218397733845

Critical Values:

1%: -3.6461350877925254

5%: -2.954126991123355

10%: -2.6159676124885216

Non-Stationary (Fail to reject the null hypothesis)

```
[ ]: numerical_columns = final_dataset.select_dtypes(include=['number'])

#calculating the mean, median and standard deviation for numerical variables
print("\n\nMean of final_dataset:\n")
print(numerical_columns.mean())

print("\n\nMedian of final_dataset:\n")
print(numerical_columns.median())

print("\n\nStandard Deviation of final_dataset:\n")
print(numerical_columns.std())
```

Mean of final_dataset:

month_name	6.248408
year_num	0.776779
shop_id	33.002959
item_id	10200.281275

```
category_id      40.016374
item_price       889.362730
item_cnt_month   1.247257
revenue          1164.267451
log_revenue      6.254676
scaled_revenue   0.000636
dtype: float64
```

Median of final_dataset:

```
month_name       6.000000
year_num         1.000000
shop_id          31.000000
item_id          9355.000000
category_id      40.000000
item_price       399.000000
item_cnt_month   1.000000
revenue          449.000000
log_revenue      6.107023
scaled_revenue   0.000245
dtype: float64
```

Standard Deviation of final_dataset:

```
month_name       3.535921
year_num         0.768460
shop_id          16.225426
item_id          6324.390884
category_id      17.098094
item_price       1718.155164
item_cnt_month   2.217429
revenue          5684.853342
log_revenue      1.171779
scaled_revenue   0.003106
dtype: float64
```

```
[ ]: print("\n\nHead of final_dataset:\n")
print(final_dataset.head(20))
print(final_dataset.shape)
print(final_dataset.info())
```

Head of final_dataset:

```
date date_num month_name year_num shop_id \

```

0	2013-01-02	02	1	0	59
1	2013-01-03	03	1	0	25
3	2013-01-06	06	1	0	25
4	2013-01-15	15	1	0	25
5	2013-01-10	10	1	0	25
6	2013-01-02	02	1	0	25
7	2013-01-04	04	1	0	25
8	2013-01-11	11	1	0	25
9	2013-01-03	03	1	0	25
10	2013-01-03	03	1	0	25
11	2013-01-05	05	1	0	25
12	2013-01-07	07	1	0	25
13	2013-01-08	08	1	0	25
14	2013-01-10	10	1	0	25
15	2013-01-11	11	1	0	25
16	2013-01-13	13	1	0	25
17	2013-01-16	16	1	0	25
18	2013-01-26	26	1	0	25
19	2013-01-27	27	1	0	25
20	2013-01-09	09	1	0	25

	shop_name	item_id	\
0	Yaroslavl TC" Altair "	22154	
1	Moscow TEC" Atrium "	2552	
3	Moscow TEC" Atrium "	2554	
4	Moscow TEC" Atrium "	2555	
5	Moscow TEC" Atrium "	2564	
6	Moscow TEC" Atrium "	2565	
7	Moscow TEC" Atrium "	2572	
8	Moscow TEC" Atrium "	2572	
9	Moscow TEC" Atrium "	2573	
10	Moscow TEC" Atrium "	2574	
11	Moscow TEC" Atrium "	2574	
12	Moscow TEC" Atrium "	2574	
13	Moscow TEC" Atrium "	2574	
14	Moscow TEC" Atrium "	2574	
15	Moscow TEC" Atrium "	2574	
16	Moscow TEC" Atrium "	2574	
17	Moscow TEC" Atrium "	2574	
18	Moscow TEC" Atrium "	2574	
19	Moscow TEC" Atrium "	2574	
20	Moscow TEC" Atrium "	2593	

	item_name	category_id	\
0	SCENE 2012 (BD)	37	
1	DEEP PURPLE The House Of Blue Light LP	58	
3	DEEP PURPLE Who Do You Think We Are LP	58	
4	DEEP PURPLE 30 Very Best Of 2CD (Businesses).	56	

5	DEEP PURPLE Perihelion: Live In Concert DVD (C...	59		
6	DEEP PURPLE Stormbringer (firms).	56		
7	DEFTONES Koi No Yakan	55		
8	DEFTONES Koi No Yakan	55		
9	DEL REY LANA Born To Die	55		
10	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
11	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
12	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
13	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
14	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
15	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
16	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
17	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
18	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
19	DEL REY LANA Born To Die The Paradise Edition 2CD	55		
20	DEPECHE MODE Music For The Masses	55		
\\				
0	item_category_name	item_price	item_cnt_month	revenue
1	Cinema - Blu-Ray	999.00	1	999.00
1	Music - Vinyl	899.00	1	899.00
3	Music - Vinyl	1709.05	1	1709.05
4	Music - CD of corporate production	1099.00	1	1099.00
5	Music - Music video	349.00	1	349.00
6	Music - CD of corporate production	549.00	1	549.00
7	Music - CD of local production	239.00	1	239.00
8	Music - CD of local production	299.00	1	299.00
9	Music - CD of local production	299.00	3	897.00
10	Music - CD of local production	399.00	2	798.00
11	Music - CD of local production	399.00	1	399.00
12	Music - CD of local production	399.00	1	399.00
13	Music - CD of local production	399.00	2	798.00
14	Music - CD of local production	399.00	1	399.00
15	Music - CD of local production	399.00	2	798.00
16	Music - CD of local production	399.00	1	399.00
17	Music - CD of local production	399.00	1	399.00
18	Music - CD of local production	399.00	1	399.00
19	Music - CD of local production	399.00	1	399.00
20	Music - CD of local production	279.00	1	279.00
\\				
0	price_range	log_revenue	scaled_revenue	
0	900-100000	6.906755	0.000546	
1	800-900	6.801283	0.000491	
3	900-100000	7.443693	0.000934	
4	900-100000	7.002156	0.000601	
5	300-400	5.855072	0.000191	
6	500-600	6.308098	0.000300	
7	200-300	5.476464	0.000131	
8	200-300	5.700444	0.000163	

```

9      200-300    6.799056    0.000490
10     300-400    6.682109    0.000436
11     300-400    5.988961    0.000218
12     300-400    5.988961    0.000218
13     300-400    6.682109    0.000436
14     300-400    5.988961    0.000218
15     300-400    6.682109    0.000436
16     300-400    5.988961    0.000218
17     300-400    5.988961    0.000218
18     300-400    5.988961    0.000218
19     300-400    5.988961    0.000218
20     200-300    5.631212    0.000152
(2928483, 16)
<class 'pandas.core.frame.DataFrame'>
Index: 2928483 entries, 0 to 2935848
Data columns (total 16 columns):
 #   Column            Dtype  
--- 
 0   date              datetime64[ns]
 1   date_num          object 
 2   month_name        int64  
 3   year_num          int64  
 4   shop_id           int64  
 5   shop_name          object 
 6   item_id            int64  
 7   item_name          object 
 8   category_id       int64  
 9   item_category_name object 
 10  item_price         float64
 11  item_cnt_month    int64  
 12  revenue            float64
 13  price_range        category
 14  log_revenue        float64
 15  scaled_revenue     float64
dtypes: category(1), datetime64[ns](1), float64(4), int64(6), object(4)
memory usage: 360.3+ MB
None

```

```
[ ]: #export the final dataset to csv file
final_dataset.to_csv('./data-set/output/final_dataset_with_cleaning.csv', index=False)
```

1.5 Model Development, Error Analysis & Comparison

```
[ ]: #prepare the data for modeling
df = pd.read_csv('./data-set/sales_train.csv')
#rename item_cnt_day column
```

```

df.rename(columns={'item_cnt_day': 'item_count'}, inplace=True)
#removes duplicates
df.drop_duplicates(inplace=True)
#outlier treatment
df = df[(df['item_count'] > 0) & (df['item_count'] < 307980)]
df = df[df['item_count'] < 1000]
#handles incorrect data
df = df[(df['item_price'] > 0) & (df['item_price'] < 100000)]
#converts date column to datetime format
df['date'] = pd.to_datetime(df['date'], format='%d.%m.%Y')
#convert date to year-month format
df['year-month'] = df['date'].dt.strftime('%Y-%m')
#drop date column and item_price column
df.drop(columns=['date', 'item_price'], inplace=True)
# group features
df_train_group = df.groupby(['year-month', 'shop_id', 'item_id']).sum()
    ↪reset_index()
# pivot table
df = df_train_group.pivot_table(index=['shop_id', 'item_id'],
    ↪columns='year-month', values='item_count', fill_value=0).reset_index()

print(df.head(10))
print(df.shape)
print(df.info())

```

year-month	shop_id	item_id	2013-01	2013-02	2013-03	2013-04	2013-05	\
0	0	30	0	31	0	0	0	
1	0	31	0	11	0	0	0	
2	0	32	6	10	0	0	0	
3	0	33	3	3	0	0	0	
4	0	35	1	14	0	0	0	
5	0	36	0	1	0	0	0	
6	0	40	0	1	0	0	0	
7	0	42	0	1	0	0	0	
8	0	43	1	0	0	0	0	
9	0	49	0	2	0	0	0	

year-month	2013-06	2013-07	2013-08	...	2015-01	2015-02	2015-03	\
0	0	0	0	...	0	0	0	
1	0	0	0	...	0	0	0	
2	0	0	0	...	0	0	0	
3	0	0	0	...	0	0	0	
4	0	0	0	...	0	0	0	
5	0	0	0	...	0	0	0	
6	0	0	0	...	0	0	0	
7	0	0	0	...	0	0	0	
8	0	0	0	...	0	0	0	

```

9          0      0      0 ...      0      0      0      0
year-month 2015-04 2015-05 2015-06 2015-07 2015-08 2015-09 2015-10
0          0      0      0 ...      0      0      0      0
1          0      0      0 ...      0      0      0      0
2          0      0      0 ...      0      0      0      0
3          0      0      0 ...      0      0      0      0
4          0      0      0 ...      0      0      0      0
5          0      0      0 ...      0      0      0      0
6          0      0      0 ...      0      0      0      0
7          0      0      0 ...      0      0      0      0
8          0      0      0 ...      0      0      0      0
9          0      0      0 ...      0      0      0      0

```

[10 rows x 36 columns]
(424097, 36)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 424097 entries, 0 to 424096
Data columns (total 36 columns):

#	Column	Non-Null Count	Dtype	
0	shop_id	424097	non-null	int64
1	item_id	424097	non-null	int64
2	2013-01	424097	non-null	int64
3	2013-02	424097	non-null	int64
4	2013-03	424097	non-null	int64
5	2013-04	424097	non-null	int64
6	2013-05	424097	non-null	int64
7	2013-06	424097	non-null	int64
8	2013-07	424097	non-null	int64
9	2013-08	424097	non-null	int64
10	2013-09	424097	non-null	int64
11	2013-10	424097	non-null	int64
12	2013-11	424097	non-null	int64
13	2013-12	424097	non-null	int64
14	2014-01	424097	non-null	int64
15	2014-02	424097	non-null	int64
16	2014-03	424097	non-null	int64
17	2014-04	424097	non-null	int64
18	2014-05	424097	non-null	int64
19	2014-06	424097	non-null	int64
20	2014-07	424097	non-null	int64
21	2014-08	424097	non-null	int64
22	2014-09	424097	non-null	int64
23	2014-10	424097	non-null	int64
24	2014-11	424097	non-null	int64
25	2014-12	424097	non-null	int64
26	2015-01	424097	non-null	int64

```
27 2015-02 424097 non-null int64
28 2015-03 424097 non-null int64
29 2015-04 424097 non-null int64
30 2015-05 424097 non-null int64
31 2015-06 424097 non-null int64
32 2015-07 424097 non-null int64
33 2015-08 424097 non-null int64
34 2015-09 424097 non-null int64
35 2015-10 424097 non-null int64
dtypes: int64(36)
memory usage: 116.5 MB
None
```

```
[ ]: #export the final dataset to csv file
df.to_csv('./data-set/output/dataset_for_modeling.csv', index=False)
```

```
[ ]: # Create X and y variables for train and test sets
```

```
X = df[df.columns[:-1]]
y = df[df.columns[-1]]
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,random_state=42)
```

```
print(X_train.shape)
print(X_test.shape)
print(y_train.shape)
print(y_test.shape)
```

```
(339277, 35)
(84820, 35)
(339277,)
(84820,)
```

```
[ ]: #creating evaluation metrics
```

```
scores_and_names = []
```

```
# Create a function to evaluate the model
def evaluate_the_model(y_true, y_pred, model_name, model, index=None):
    if not isinstance(y_true, pd.Series):
        if index is None:
            index = range(len(y_true))
        y_true = pd.Series(y_true, index=index)

    # Calculate the MAE
    mae = mean_absolute_error(y_true, y_pred)
    print(f"MAE for {model_name}: {mae:.5f}")
```

```

# Calculate the MSE
mse = mean_squared_error(y_true, y_pred)
print(f"MSE for {model_name}: {mse:.5f}")

# Calculate the RMSE
rmse = np.sqrt(mse)
print(f"RMSE for {model_name}: {rmse:.5f}")

# calculate r2 score
r2 = r2_score(y_true, y_pred)
print(f"R2 for {model_name}: {r2:.5f}")

# Plot the predictions vs. the actual values
fig, ax = plt.subplots(figsize=(12, 6))
sns.lineplot(x=y_true.index, y=y_true, label='Actual Values')
sns.lineplot(x=y_true.index, y=y_pred, label='Predicted Values')
plt.title(f'Predictions vs. Actual Values ({model_name})')
plt.xlabel('Observation')
plt.ylabel('Item Count')
plt.legend()
plt.show()

# # Plot the predictions vs. the actual values
# plt.figure(figsize=(12, 6))
# plt.plot(y_true, label='Actual Values')
# plt.plot(y_pred, label='Predicted Values')
# plt.title(f'Predictions vs. Actual Values ({model_name})')
# plt.xlabel('Observation')
# plt.ylabel('Item Count')
# plt.legend()
# plt.show()

scores_and_names.append((model_name, rmse))

```

1.5.1 Decision Tree

```

[ ]: # create a decision tree model
dt = DecisionTreeRegressor(random_state=42)
dt.fit(X_train, y_train)
y_pred = dt.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Decision Tree', dt)

#visualize the decision tree
feature_names = list(X.columns)

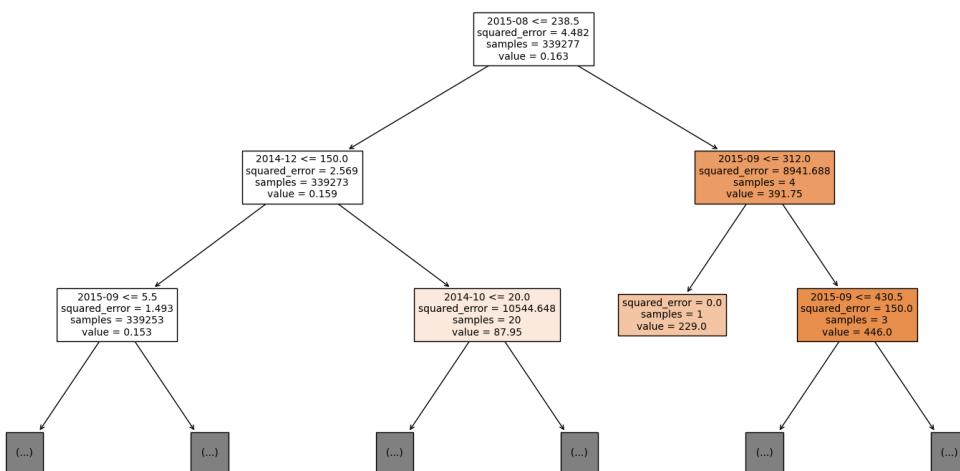
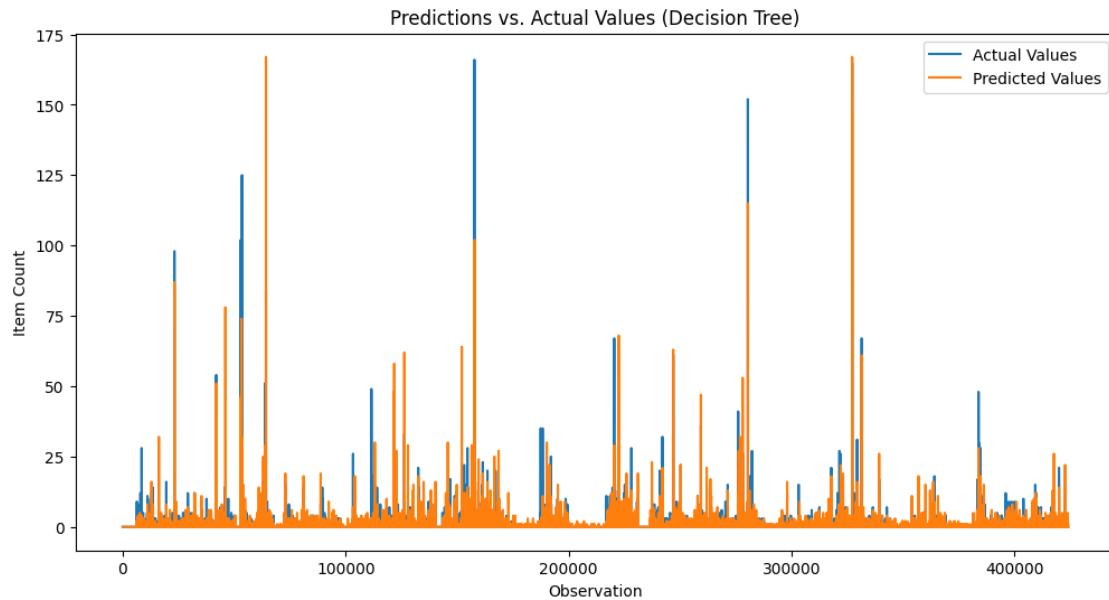
```

```

plt.figure(figsize=(20, 10))
plot_tree(dt, filled=True, feature_names=feature_names, max_depth=2,
          fontsize=10)
plt.show()

```

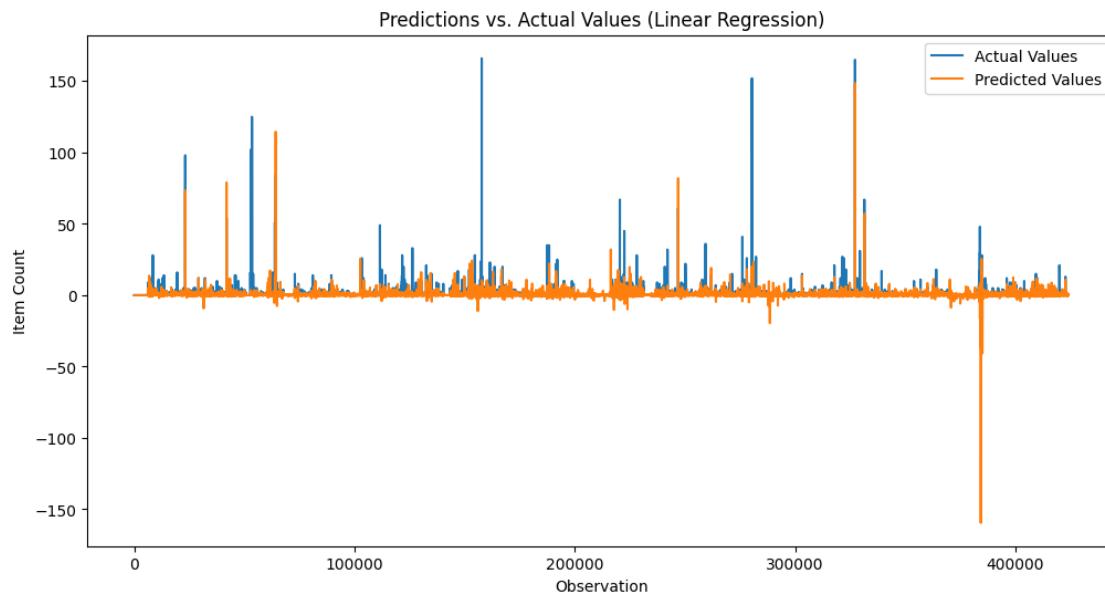
MAE for Decision Tree: 0.17993
 MSE for Decision Tree: 1.39769
 RMSE for Decision Tree: 1.18224
 R2 for Decision Tree: 0.46010



1.5.2 linear regression

```
[ ]: # Create a Linear Regression model  
lin_reg = LinearRegression()  
lin_reg.fit(X_train, y_train)  
y_pred = lin_reg.predict(X_test)  
  
evaluate_the_model(y_test, y_pred, 'Linear Regression', lin_reg)
```

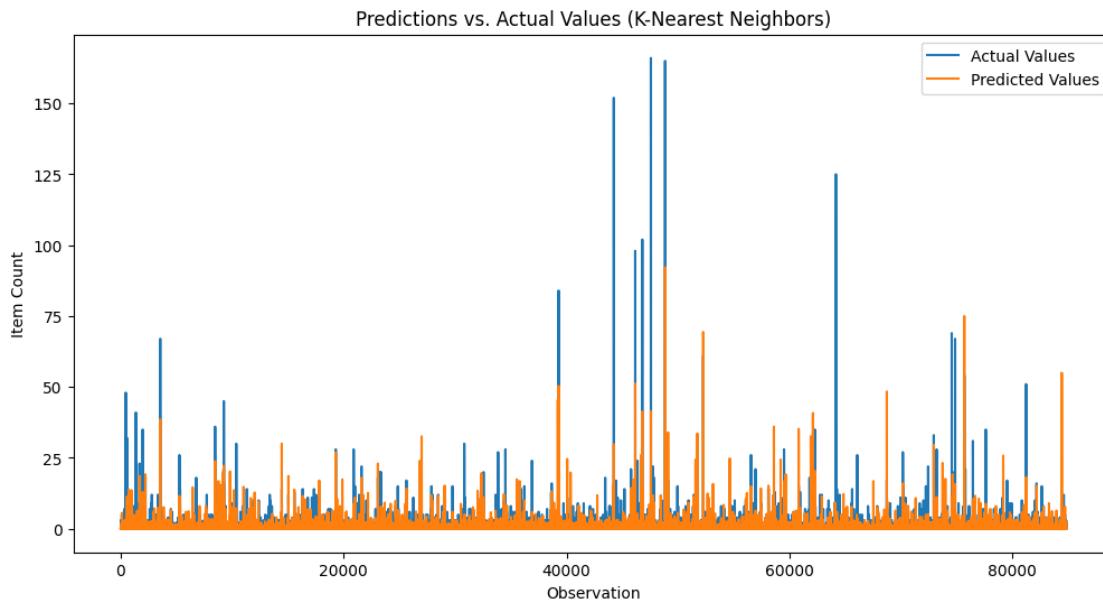
MAE for Linear Regression: 0.29022
MSE for Linear Regression: 2.39626
RMSE for Linear Regression: 1.54798
R2 for Linear Regression: 0.07438



1.5.3 KNN

```
[ ]: # create a knn model  
knn = KNeighborsRegressor(n_neighbors=5)  
knn.fit(X_train.values, y_train.values)  
y_pred = knn.predict(X_test.values)  
  
evaluate_the_model(y_test.values, y_pred, 'K-Nearest Neighbors', knn)
```

MAE for K-Nearest Neighbors: 0.18599
MSE for K-Nearest Neighbors: 1.60036
RMSE for K-Nearest Neighbors: 1.26505
R2 for K-Nearest Neighbors: 0.38182

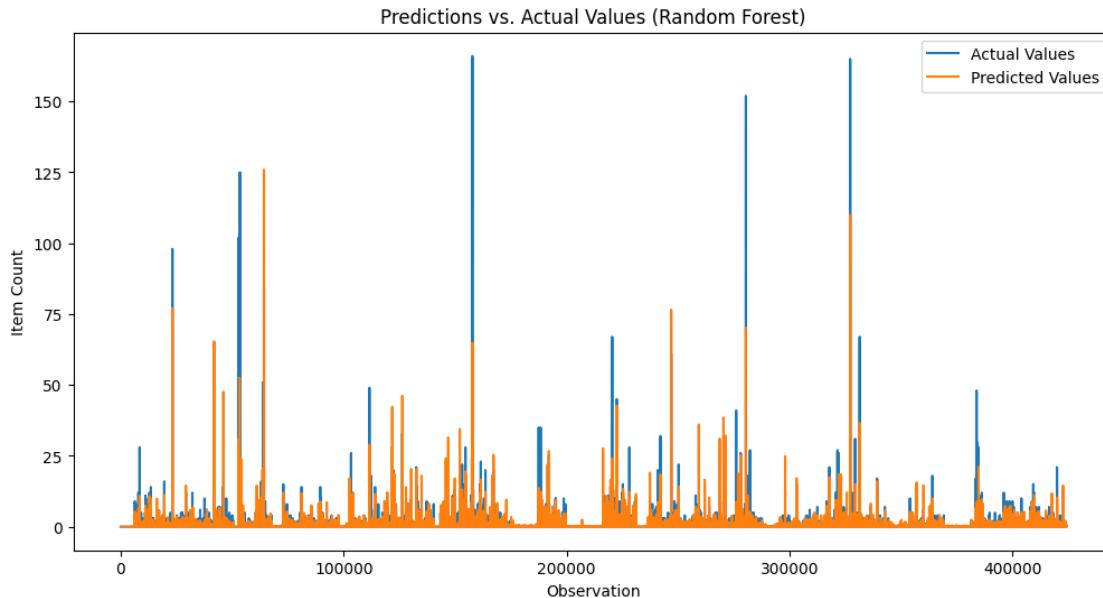


1.5.4 random forest

```
[ ]: # create a random forest model
rf = RandomForestRegressor()
rf.fit(X_train, y_train)
y_pred = rf.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Random Forest', rf)
```

MAE for Random Forest: 0.16271
 MSE for Random Forest: 1.13705
 RMSE for Random Forest: 1.06633
 R2 for Random Forest: 0.56078



1.5.5 Logistic Regression

```
[ ]: # Create a logistic regression model
log_reg = LogisticRegression()
log_reg.fit(X_train, y_train)
y_pred = log_reg.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Logistic Regression', log_reg)
```

```
C:\Users\srum\AppData\Roaming\Python\Python311\site-
packages\sklearn\linear_model\_logistic.py:460: ConvergenceWarning: lbfgs failed
to converge (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
```

Increase the number of iterations (max_iter) or scale the data as shown in:

<https://scikit-learn.org/stable/modules/preprocessing.html>

Please also refer to the documentation for alternative solver options:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

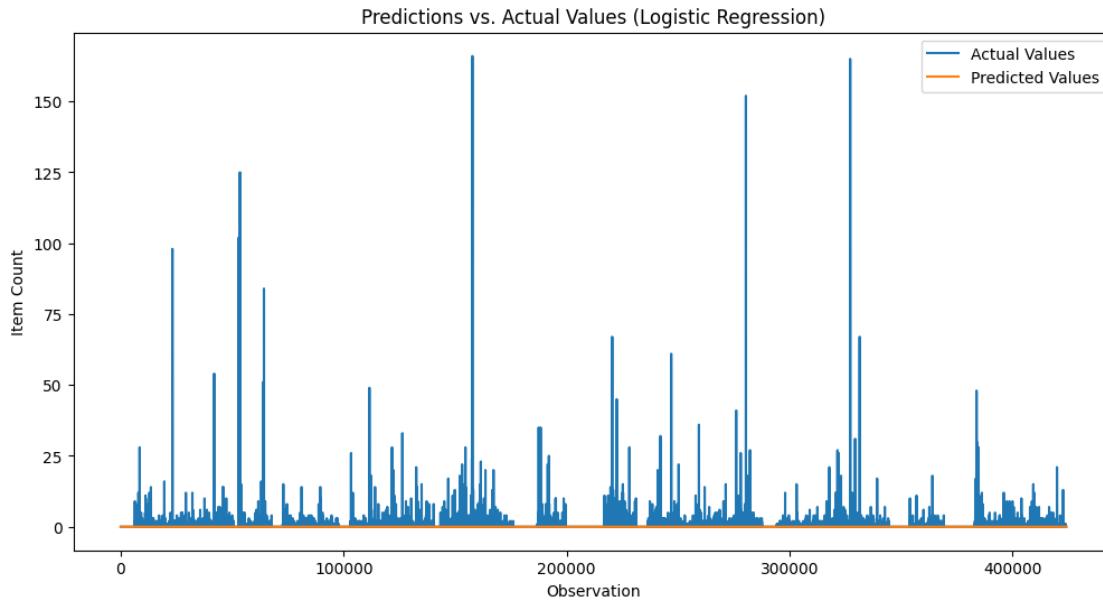
```
n_iter_i = _check_optimize_result(
```

MAE for Logistic Regression: 0.16065

MSE for Logistic Regression: 2.61462

RMSE for Logistic Regression: 1.61698

R2 for Logistic Regression: -0.00997

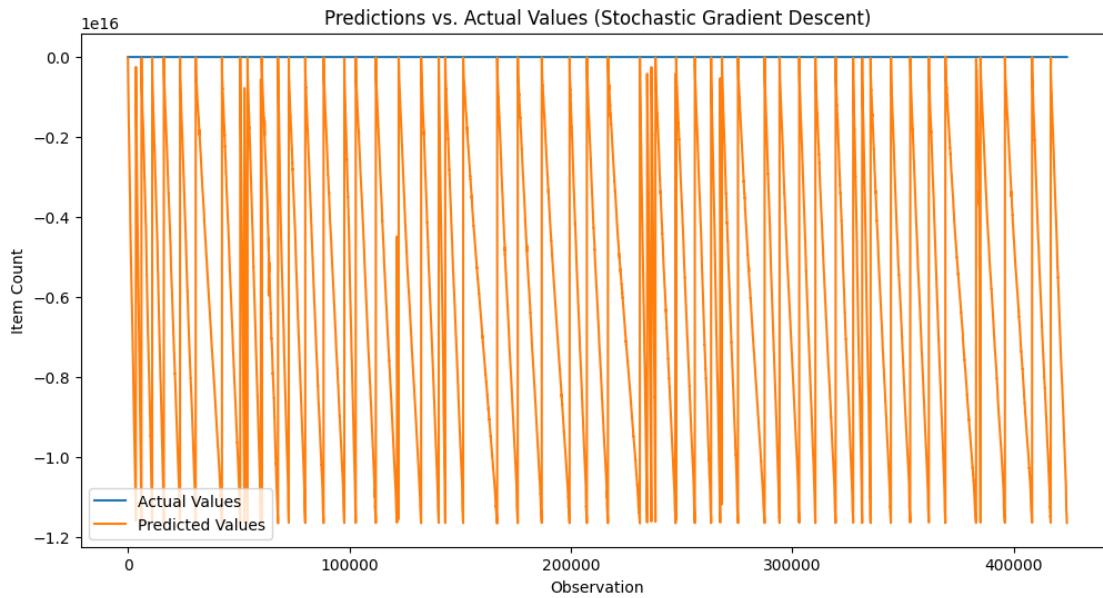


1.5.6 Stochastic Gradient Descent

```
[ ]: #create a stochastic gradient descent model
sgd_reg = SGDRegressor()
sgd_reg.fit(X_train, y_train)
y_pred = sgd_reg.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Stochastic Gradient Descent', sgd_reg)
```

MAE for Stochastic Gradient Descent: 6024432560571046.00000
MSE for Stochastic Gradient Descent: 46688412185564450841545509175296.00000
RMSE for Stochastic Gradient Descent: 6832891934281154.00000
R2 for Stochastic Gradient Descent: -18034686025774622267697019224064.00000

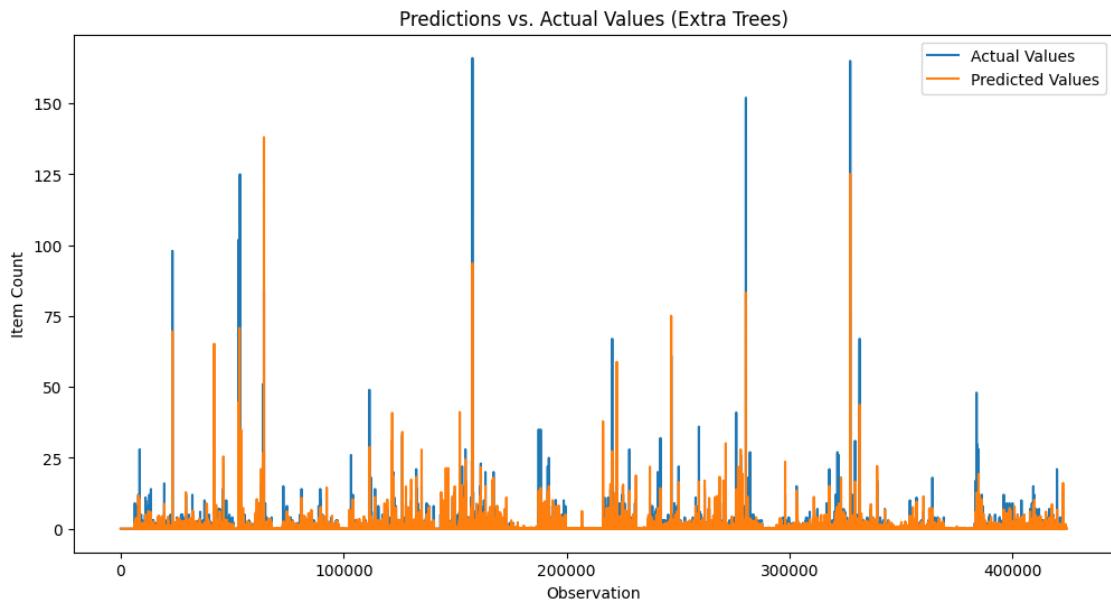


1.5.7 xtra tree

```
[ ]: #create a extra trees model
et = ExtraTreesRegressor()
et.fit(X_train, y_train)
y_pred = et.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Extra Trees', et)
```

MAE for Extra Trees: 0.17622
MSE for Extra Trees: 1.10411
RMSE for Extra Trees: 1.05077
R2 for Extra Trees: 0.57351

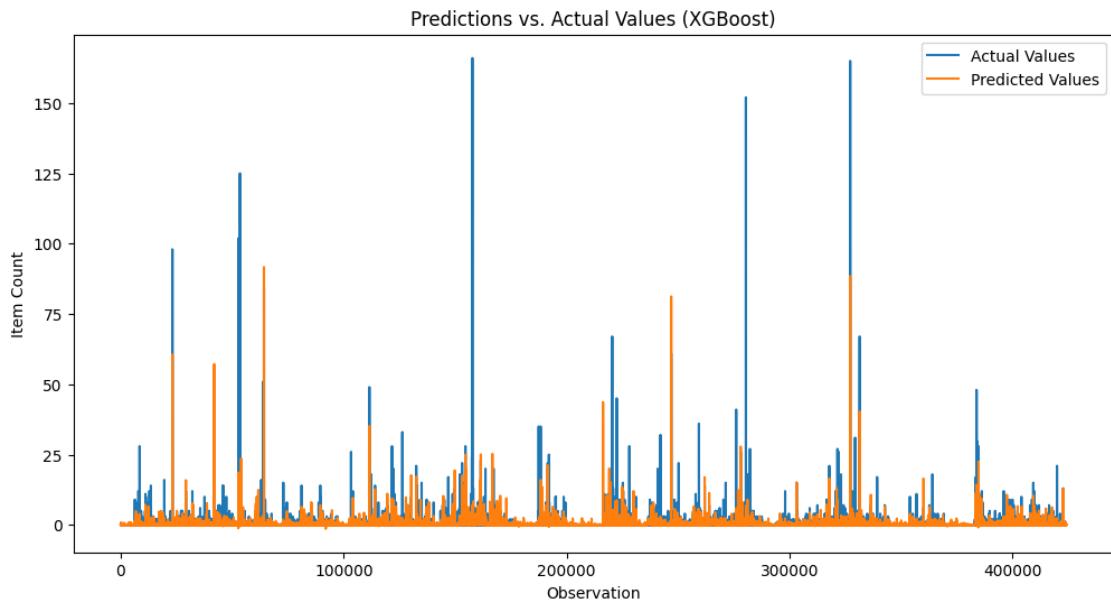


1.5.8 XGBoost

```
[ ]: #create a xgboost model
xgb = XGBRegressor()
xgb.fit(X_train, y_train)
y_pred = xgb.predict(X_test)

evaluate_the_model(y_test, y_pred, 'XGBoost', xgb)
```

MAE for XGBoost: 0.21046
 MSE for XGBoost: 1.69069
 RMSE for XGBoost: 1.30027
 R2 for XGBoost: 0.34692

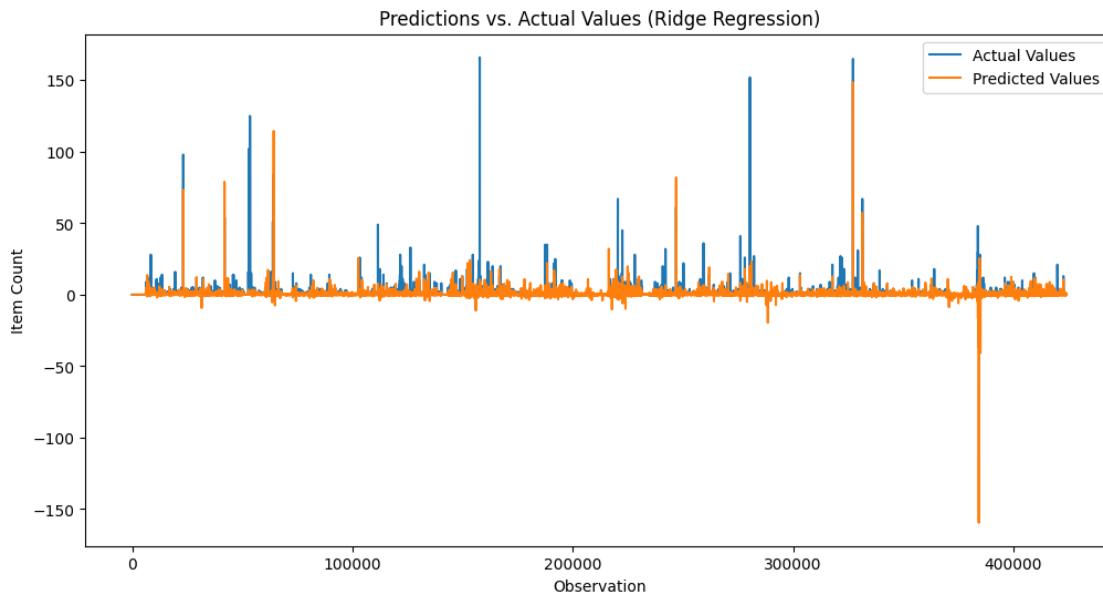


1.5.9 ridge regression

```
[ ]: #create ridge regression model
ridge = Ridge()
ridge.fit(X_train, y_train)
y_pred = ridge.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Ridge Regression', ridge)
```

MAE for Ridge Regression: 0.29022
MSE for Ridge Regression: 2.39625
RMSE for Ridge Regression: 1.54798
R2 for Ridge Regression: 0.07438

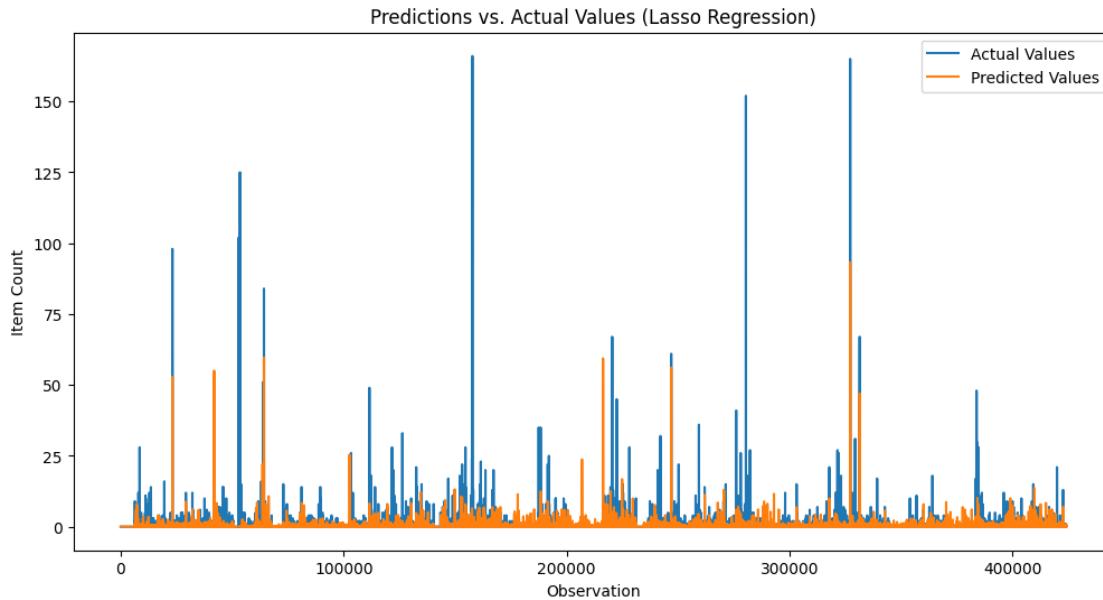


1.5.10 lasso regression

```
[ ]: #create lasso regression model
lasso = Lasso()
lasso.fit(X_train, y_train)
y_pred = lasso.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Lasso Regression', lasso)
```

MAE for Lasso Regression: 0.26487
 MSE for Lasso Regression: 2.11585
 RMSE for Lasso Regression: 1.45460
 R2 for Lasso Regression: 0.18269



1.5.11 ARIMA

```
[ ]: #create ARIMA model
arima = ARIMA(y_train, order=(1, 1, 1))
model = arima.fit()
y_pred = model.predict(start=len(y_train), end=len(y_train) + len(X_test) - 1,
exog=X_test)

evaluate_the_model(y_test, y_pred, 'ARIMA', arima)

c:\Users\srum\anaconda3\Lib\site-
packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: An unsupported
index was provided and will be ignored when e.g. forecasting.
    self._init_dates(dates, freq)
c:\Users\srum\anaconda3\Lib\site-
packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: An unsupported
index was provided and will be ignored when e.g. forecasting.
    self._init_dates(dates, freq)
c:\Users\srum\anaconda3\Lib\site-
packages\statsmodels\tsa\base\tsa_model.py:473: ValueWarning: An unsupported
index was provided and will be ignored when e.g. forecasting.
    self._init_dates(dates, freq)
c:\Users\srum\anaconda3\Lib\site-
packages\statsmodels\tsa\base\tsa_model.py:836: ValueWarning: No supported index
is available. Prediction results will be given with an integer index beginning
at `start`.
    return get_prediction_index()
```

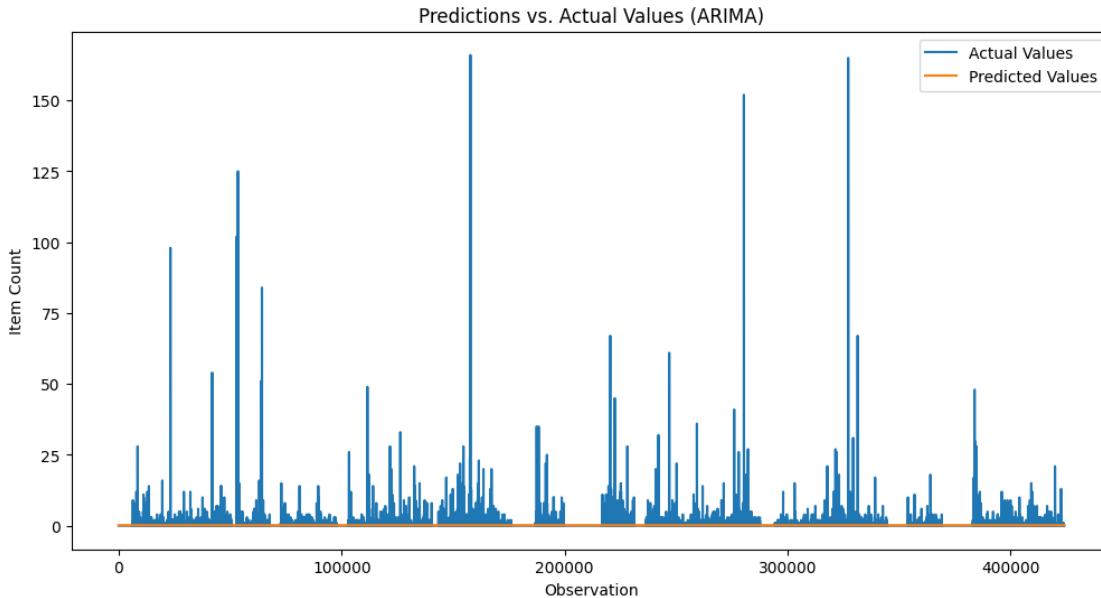
```

c:\Users\srum\anaconda3\Lib\site-
packages\statsmodels\tsa\base\tsa_model.py:836: FutureWarning: No supported
index is available. In the next version, calling this method in a model without
a supported index will result in an exception.

    return get_prediction_index()

MAE for ARIMA: 0.30035
MSE for ARIMA: 2.58882
RMSE for ARIMA: 1.60898
R2 for ARIMA: -0.00000

```



1.5.12 ADABOOST

```

[ ]: #create adaboost model
ada = AdaBoostRegressor()
ada.fit(X_train, y_train)
y_pred = ada.predict(X_test)

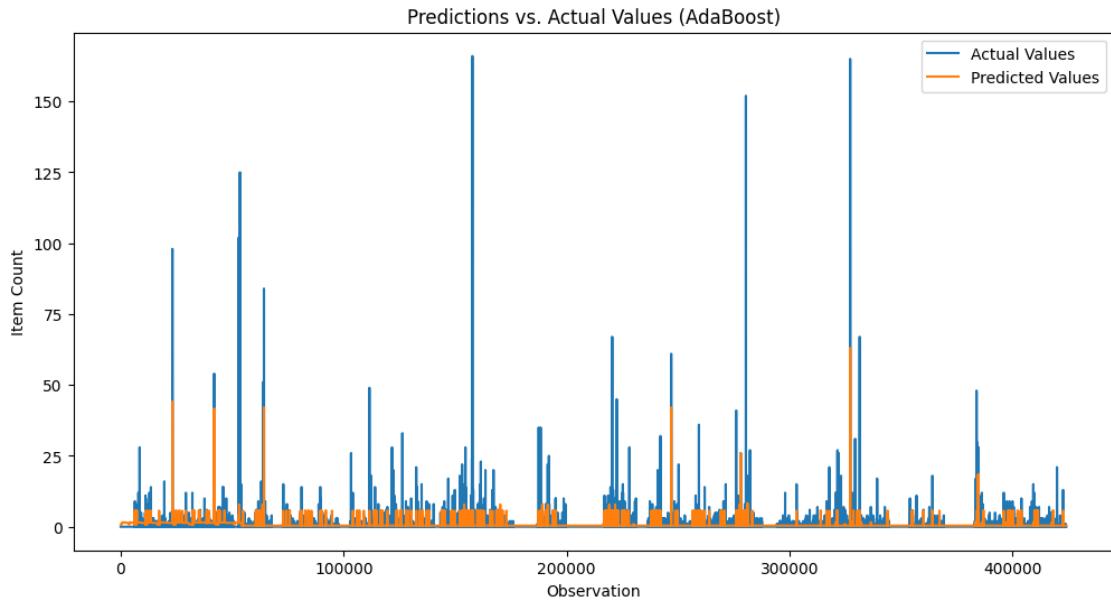
evaluate_the_model(y_test, y_pred, 'AdaBoost', ada)

```

```

MAE for AdaBoost: 0.61162
MSE for AdaBoost: 2.32629
RMSE for AdaBoost: 1.52522
R2 for AdaBoost: 0.10141

```

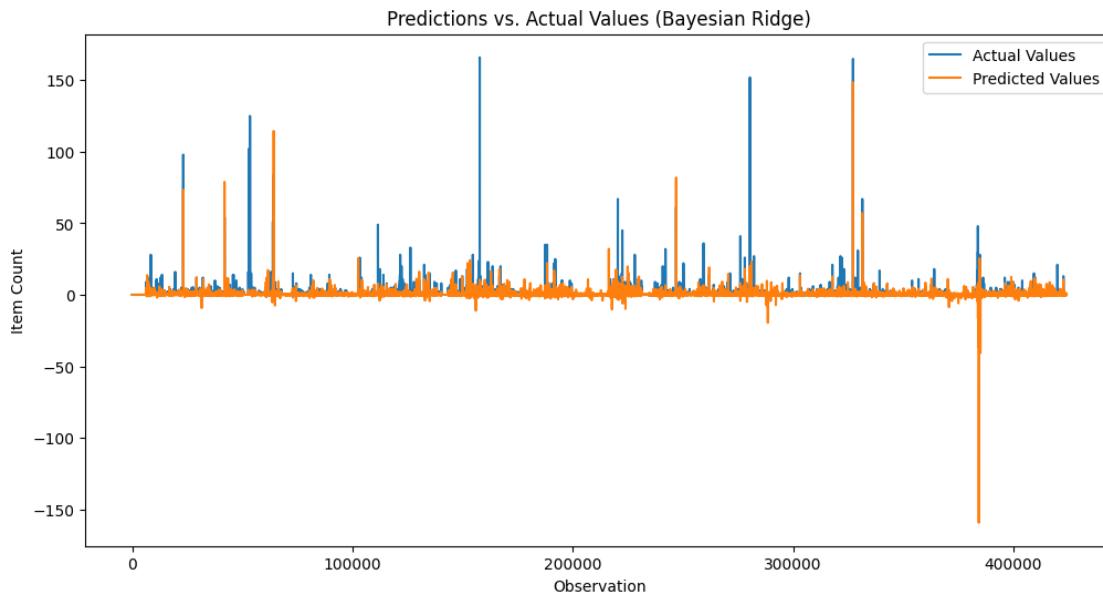


1.5.13 BayesianRidge

```
[ ]: # create bayesian ridge model
br = BayesianRidge()
br.fit(X_train, y_train)
y_pred = br.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Bayesian Ridge', br)
```

MAE for Bayesian Ridge: 0.29018
MSE for Bayesian Ridge: 2.39567
RMSE for Bayesian Ridge: 1.54780
R2 for Bayesian Ridge: 0.07461

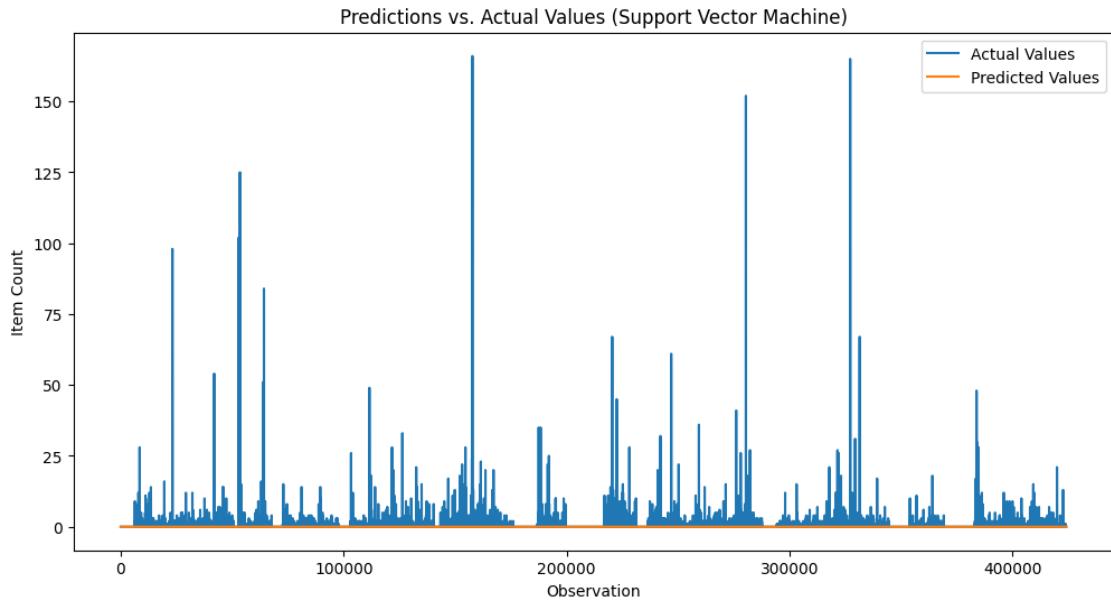


1.5.14 SVM

```
[ ]: # create a support vector machine model
svm = SVC()
svm.fit(X_train, y_train)
y_pred = svm.predict(X_test)

evaluate_the_model(y_test, y_pred, 'Support Vector Machine', svm)
```

MAE for Support Vector Machine: 0.16065
MSE for Support Vector Machine: 2.61462
RMSE for Support Vector Machine: 1.61698
R2 for Support Vector Machine: -0.00997



1.5.15 Compare Models

```
[ ]: results = pd.DataFrame(scores_and_names, columns=['Model', 'RMSE'])
```

```
[ ]: #sort the results
results.sort_values(by='RMSE', ascending=True, inplace=True)
```

```
[ ]: #print the results in tabel format
print(results)
```

	Model	RMSE
6	Extra Trees	1.050768e+00
3	Random Forest	1.066325e+00
0	Decision Tree	1.182239e+00
2	K-Nearest Neighbors	1.265051e+00
7	XGBoost	1.300267e+00
9	Lasso Regression	1.454596e+00
11	AdaBoost	1.525217e+00
12	Bayesian Ridge	1.547795e+00
8	Ridge Regression	1.547984e+00
1	Linear Regression	1.547985e+00
10	ARIMA	1.608982e+00
4	Logistic Regression	1.616978e+00
13	Support Vector Machine	1.616980e+00
5	Stochastic Gradient Descent	6.832892e+15

```
[ ]: #print the best model from the results with model name and score
print(f"\nBest Model: {results.iloc[0, 0]}")
print(f"RMSE: {results.iloc[0, 1]:.5f}")
```

Best Model: Extra Trees

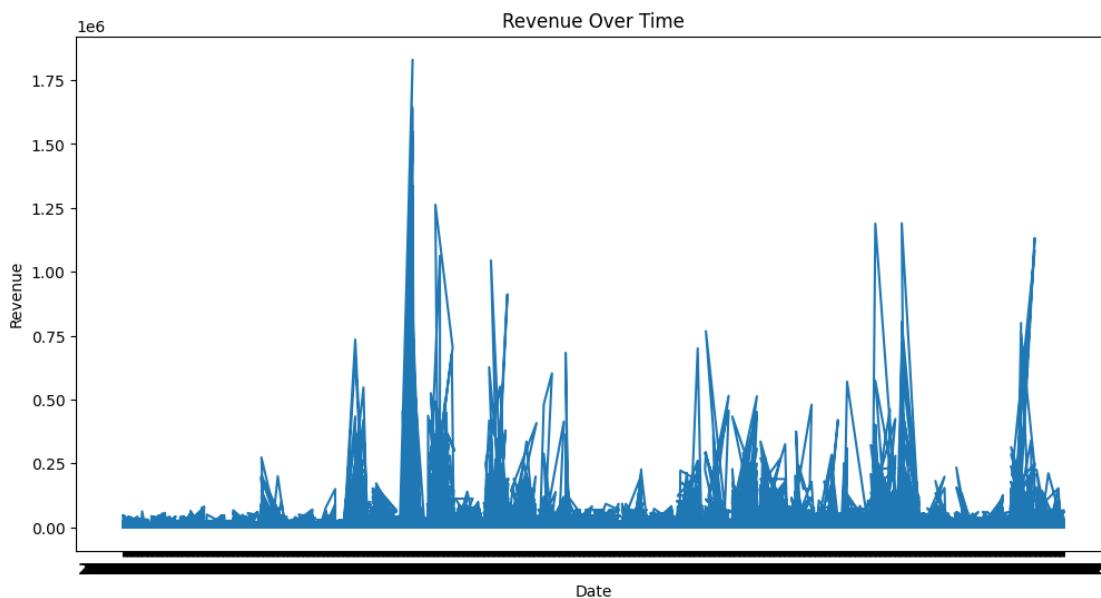
RMSE: 1.05077

1.6 Data Visualization

```
[ ]: #export the final dataset to csv file
final_dataset.to_csv('./dashboard_dataset.csv', index=False)
```

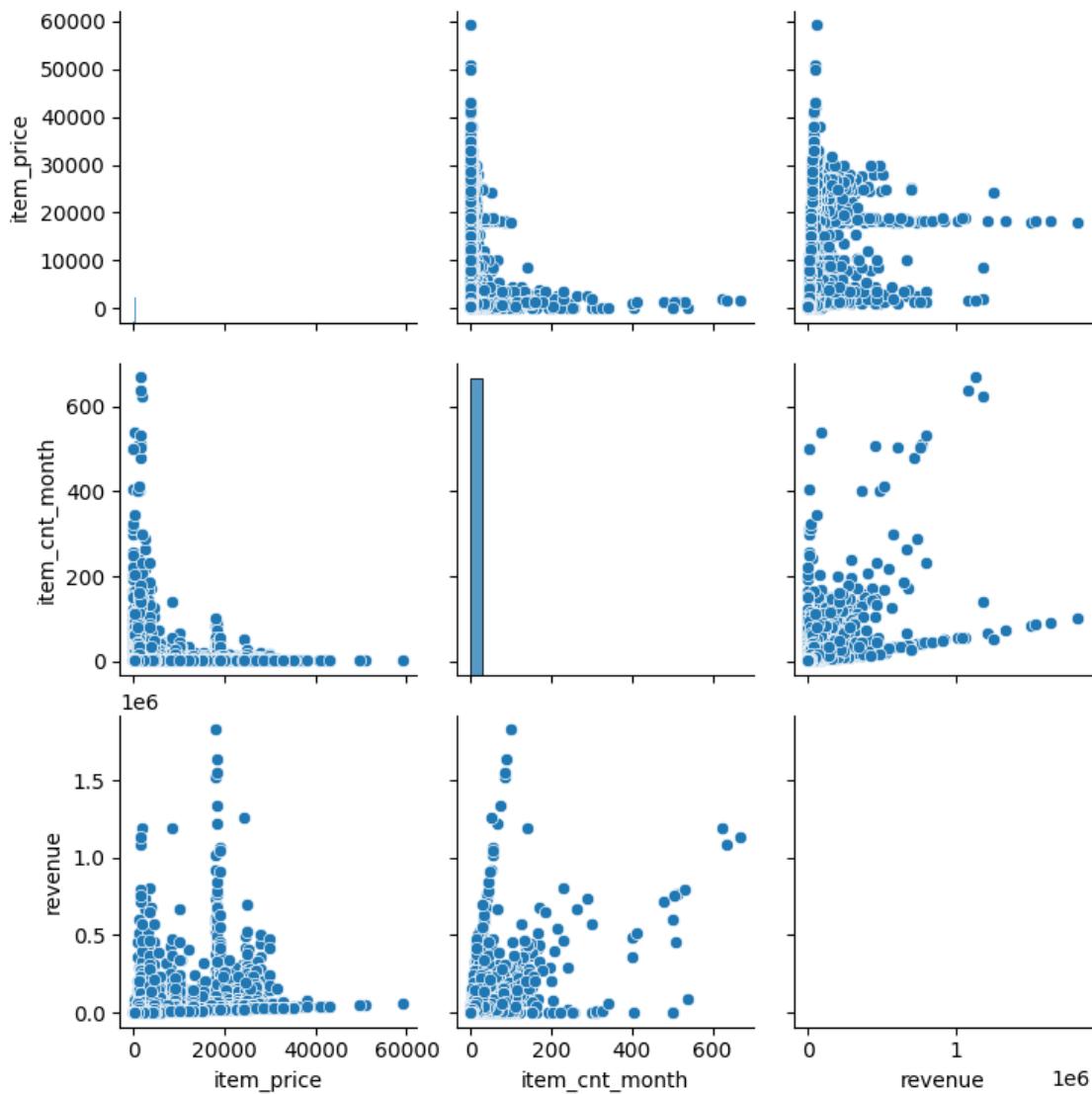
```
[ ]: #import the final dataset
final_dataset = pd.read_csv('./dashboard_dataset.csv')
```

```
[ ]: #line chart
plt.figure(figsize=(12, 6))
plt.plot(final_dataset['date'], final_dataset['revenue'])
plt.title('Revenue Over Time')
plt.xlabel('Date')
plt.ylabel('Revenue')
plt.show()
```

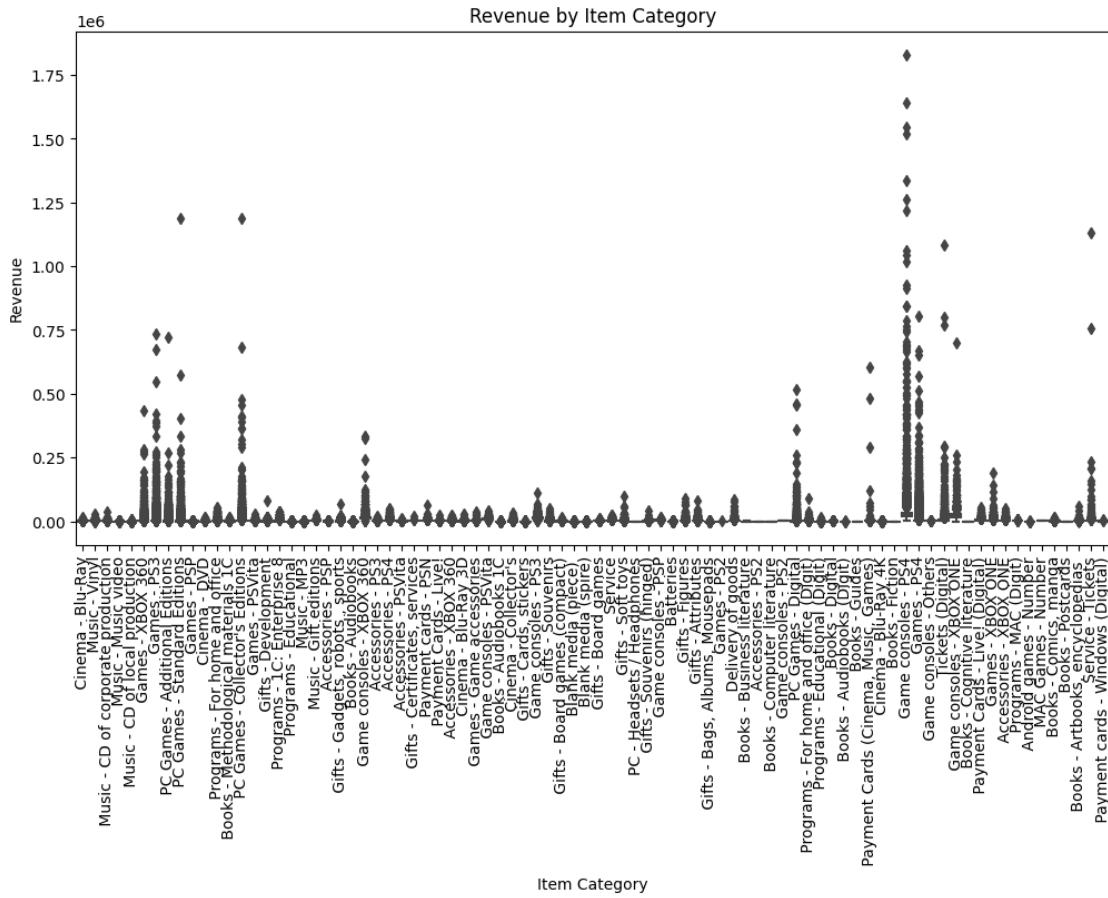


```
[ ]: #pairplot
sns.pairplot(final_dataset[['item_price', 'item_cnt_month', 'revenue']])
plt.show()
```

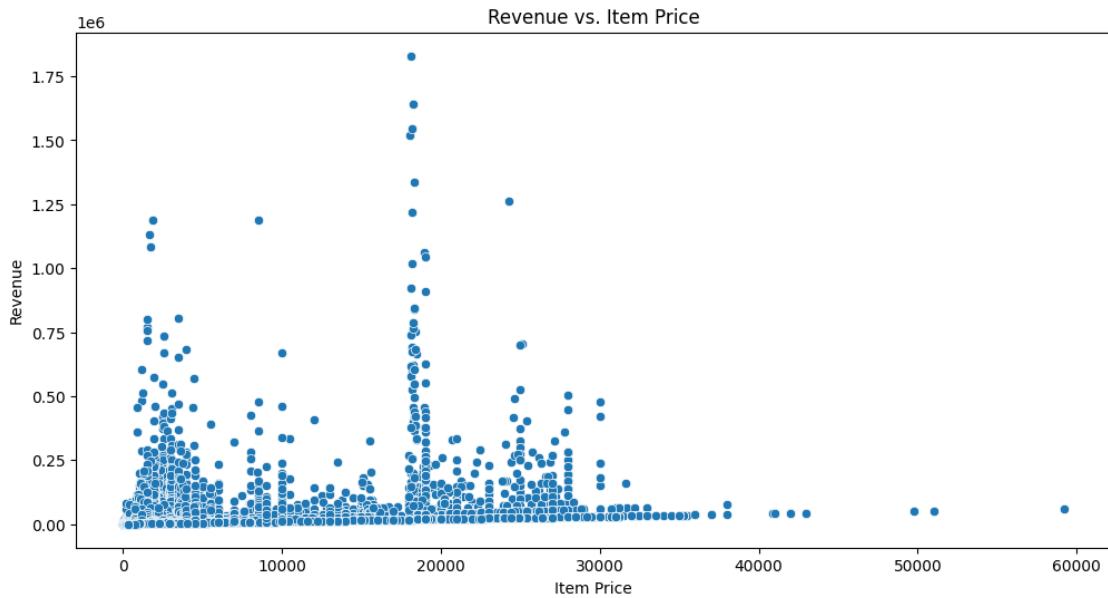
```
c:\Users\srumi\anaconda3\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning:  
The figure layout has changed to tight  
self._figure.tight_layout(*args, **kwargs)
```



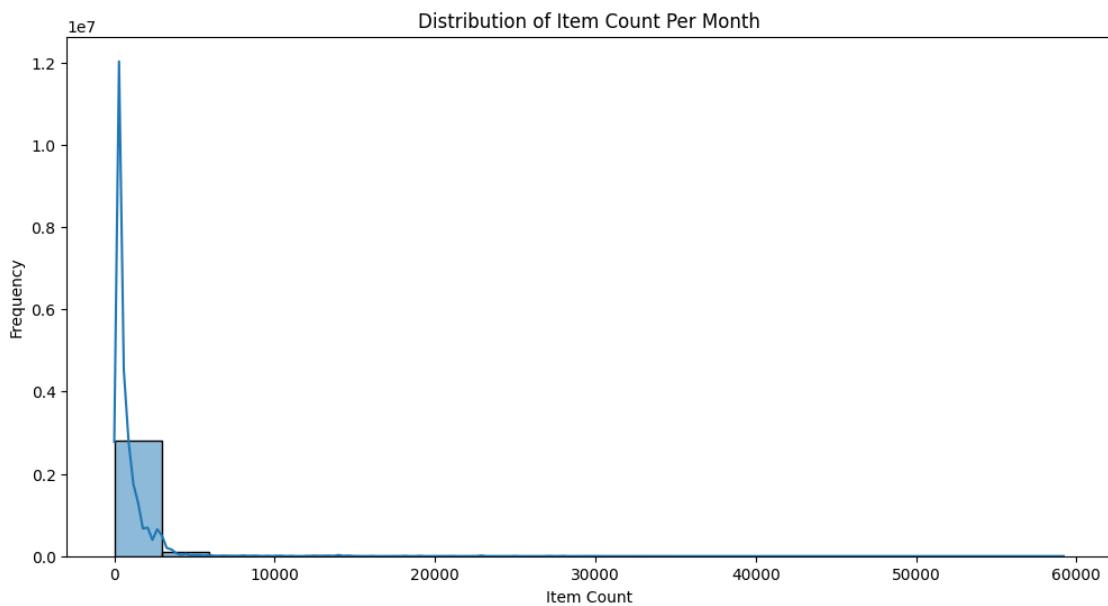
```
[ ]: #boxplot  
plt.figure(figsize=(12, 6))  
sns.boxplot(data=final_dataset, x='item_category_name', y='revenue')  
plt.title('Revenue by Item Category')  
plt.xlabel('Item Category')  
plt.ylabel('Revenue')  
plt.xticks(rotation=90)  
plt.show()
```



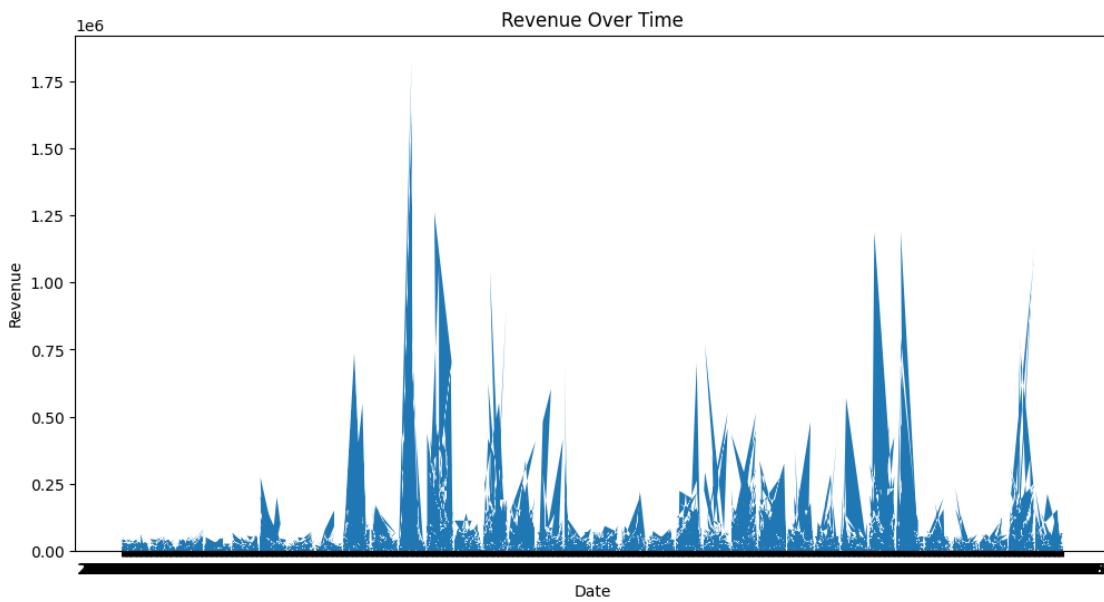
```
[ ]: #scatter chart
plt.figure(figsize=(12, 6))
sns.scatterplot(data=final_dataset, x='item_price', y='revenue')
plt.title('Revenue vs. Item Price')
plt.xlabel('Item Price')
plt.ylabel('Revenue')
plt.show()
```



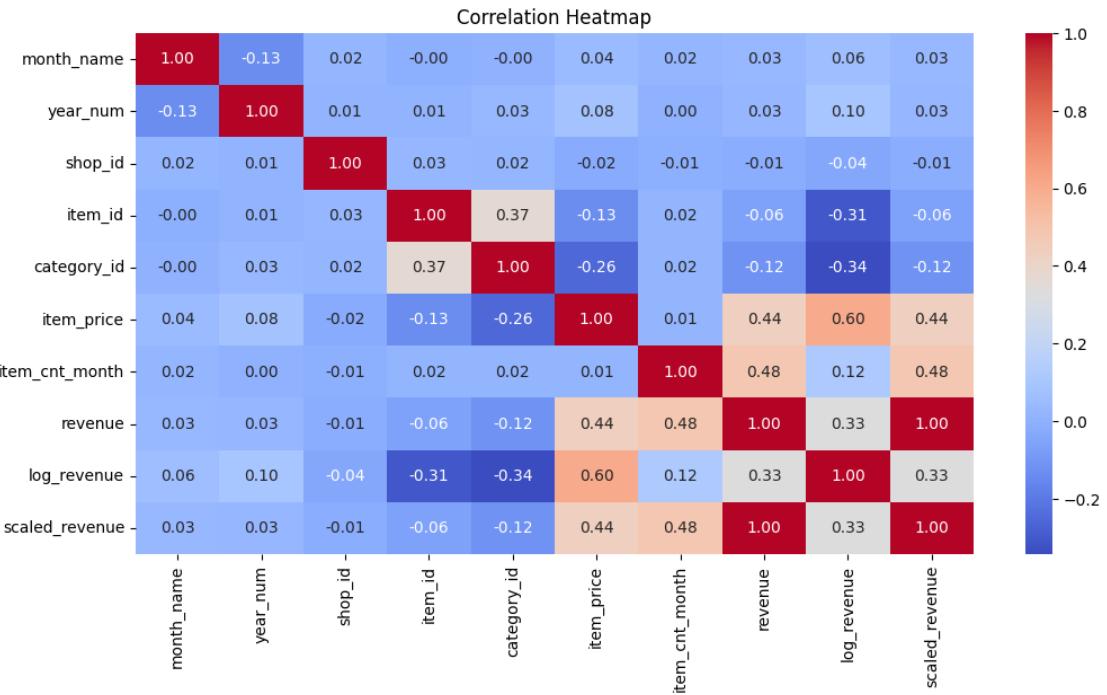
```
[ ]: #histogram
plt.figure(figsize=(12, 6))
sns.histplot(data=final_dataset, x='item_price', kde=True, bins=20)
plt.title('Distribution of Item Count Per Month')
plt.xlabel('Item Count')
plt.ylabel('Frequency')
plt.show()
```



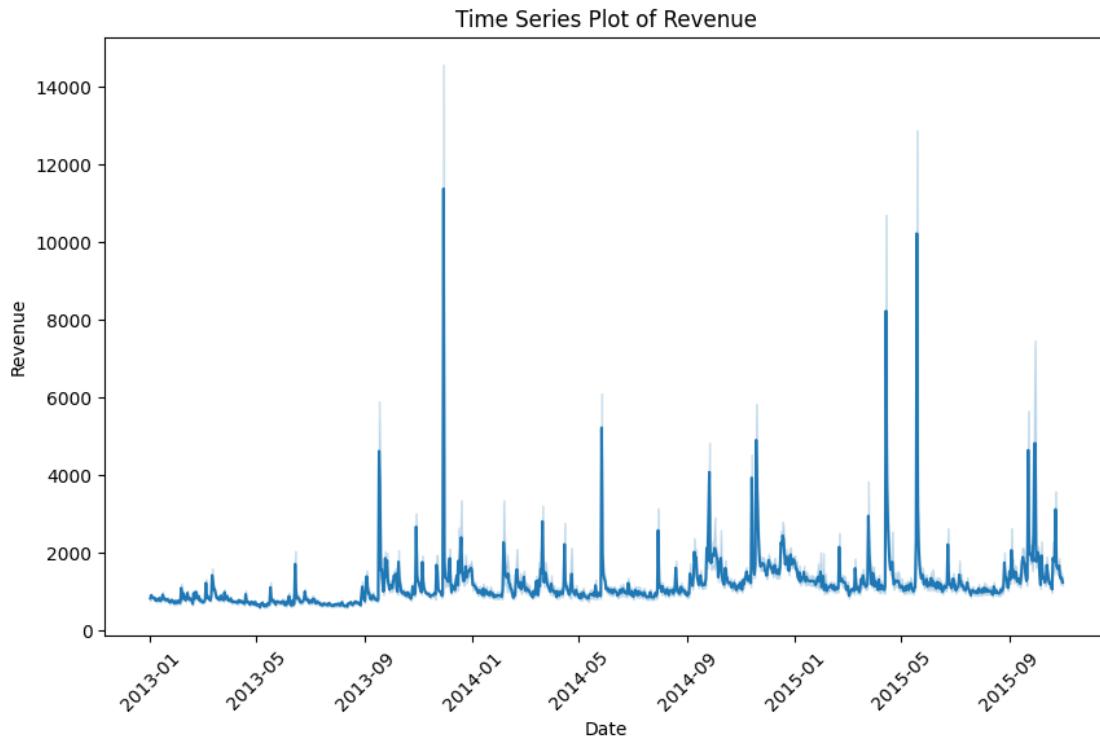
```
[ ]: #area plot
plt.figure(figsize=(12, 6))
plt.stackplot(final_dataset['date'], final_dataset['revenue'])
plt.title('Revenue Over Time')
plt.xlabel('Date')
plt.ylabel('Revenue')
plt.show()
```



```
[ ]: #heatmap
plt.figure(figsize=(12, 6))
sns.heatmap(numeric_columns.corr(), annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```

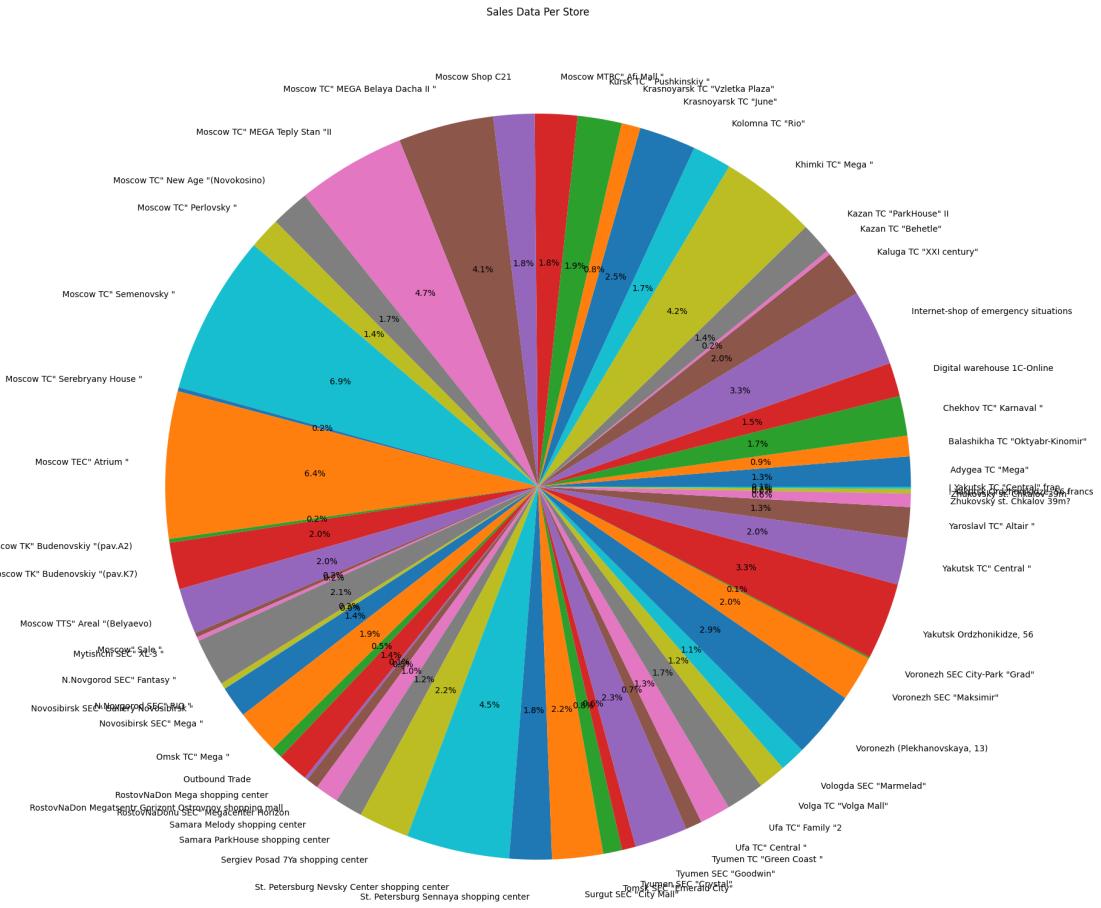


```
[ ]: # Time Series Plot
plt.figure(figsize=(10, 6))
final_dataset['date'] = pd.to_datetime(final_dataset['date'])
sns.lineplot(x='date', y='revenue', data=final_dataset)
plt.title('Time Series Plot of Revenue')
plt.xlabel('Date')
plt.ylabel('Revenue')
plt.xticks(rotation=45)
plt.show()
```

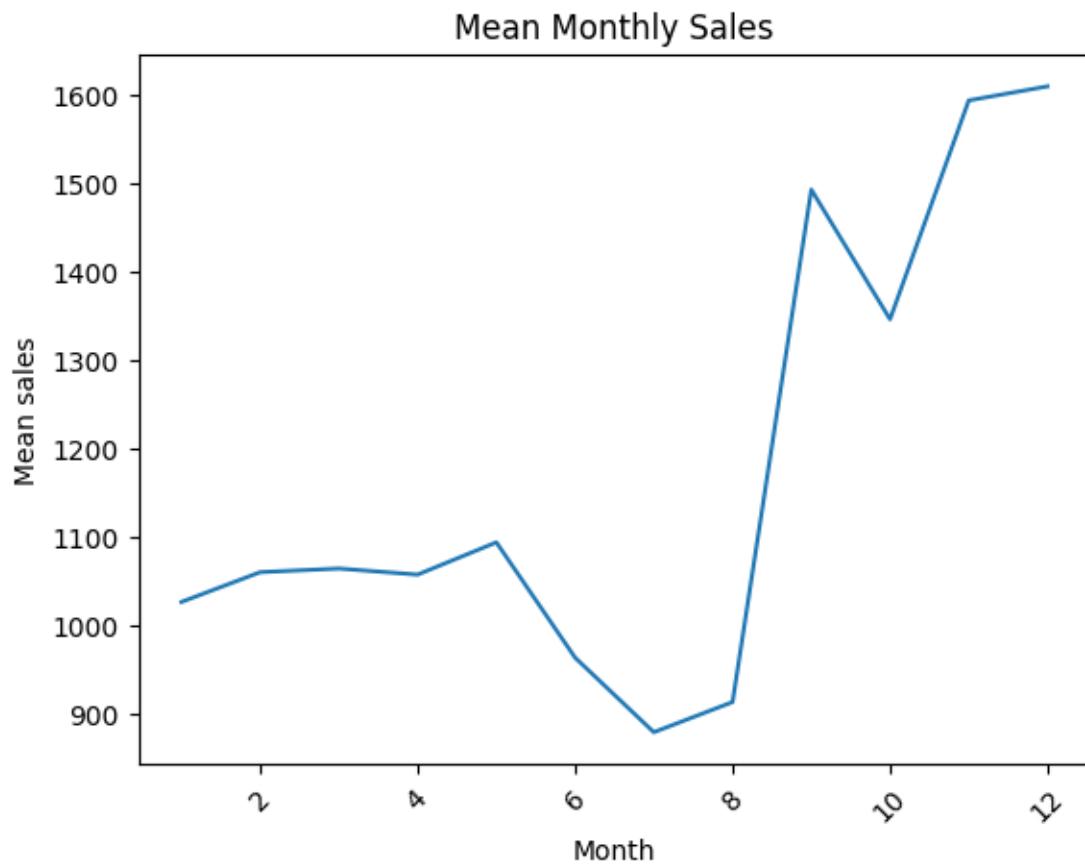


```
[ ]: #sales data per store (pie chart)
```

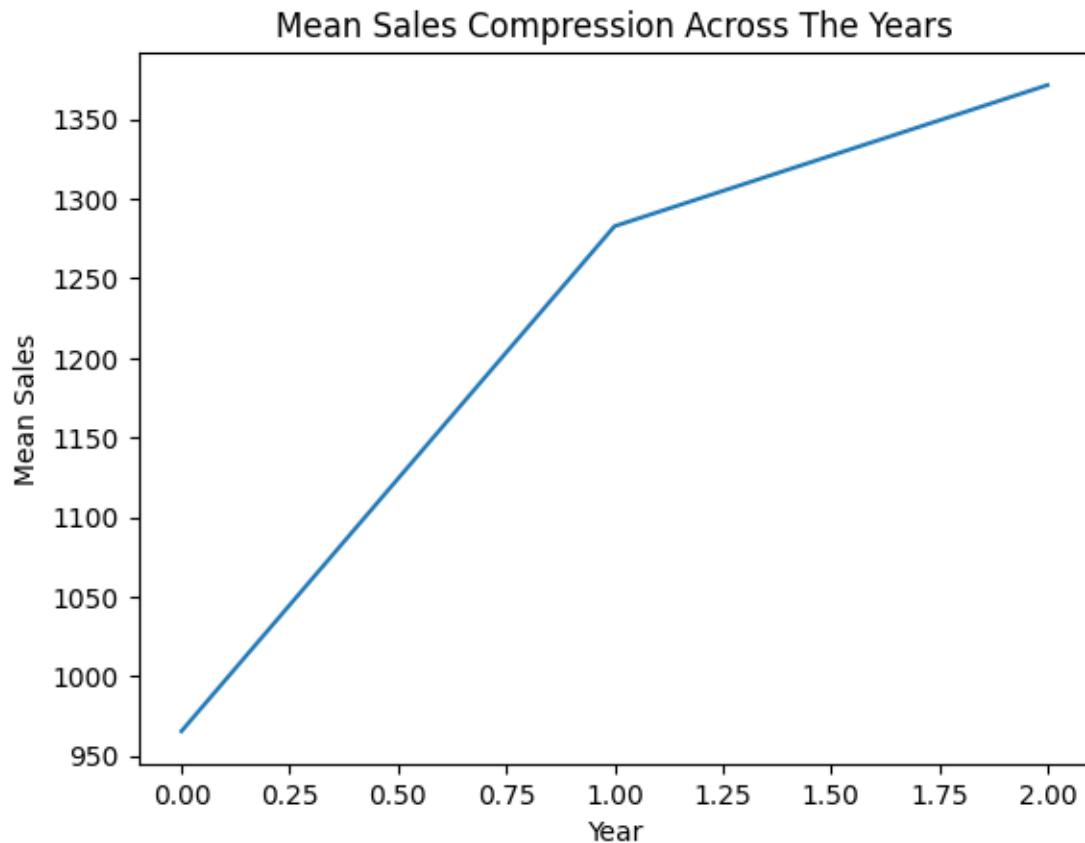
```
#group the data by shop_name and sum the revenue column
grouped_by_shop_name = final_dataset.groupby(['shop_name']).agg({'revenue': ▾
    ↵'sum'})
plt.figure(figsize=(20, 20))
plt.pie(grouped_by_shop_name['revenue'], labels=grouped_by_shop_name.index, ▾
    ↵autopct='%.1f%%')
plt.title('Sales Data Per Store')
plt.show()
```



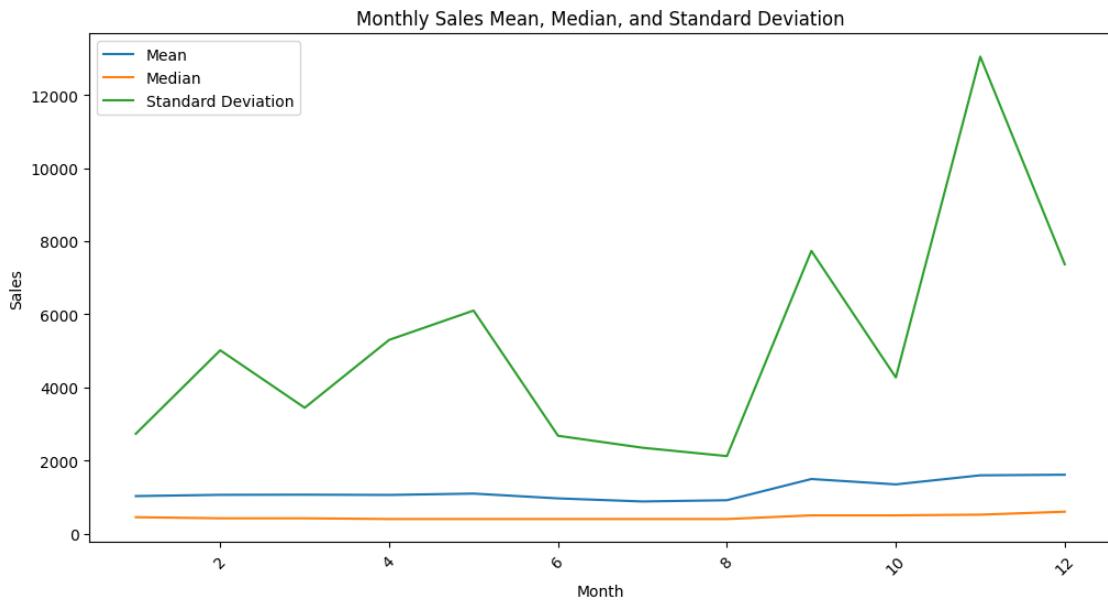
```
[ ]: #mean monthly sales
grouped_by_month_name = final_dataset.groupby(['month_name']).agg({'revenue': lambda x: x.mean()})
sns.lineplot(x=grouped_by_month_name.index, y=grouped_by_month_name['revenue'])
plt.title('Mean Monthly Sales')
plt.xlabel('Month')
plt.ylabel('Mean sales')
plt.xticks(rotation=45)
plt.show()
```



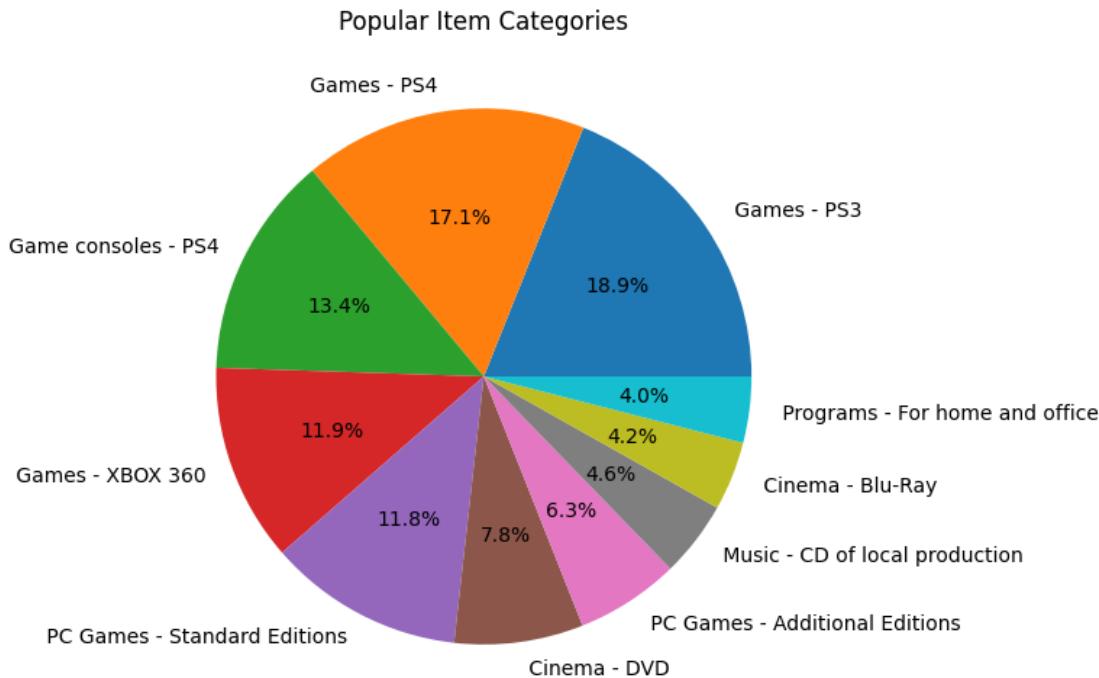
```
[ ]: #mean sales compression across the years
grouped_by_year_num = final_dataset.groupby(['year_num']).agg({'revenue': 'mean'})
sns.lineplot(x=grouped_by_year_num.index, y=grouped_by_year_num['revenue'])
plt.title('Mean Sales Compression Across The Years')
plt.xlabel('Year')
plt.ylabel('Mean Sales')
plt.show()
```



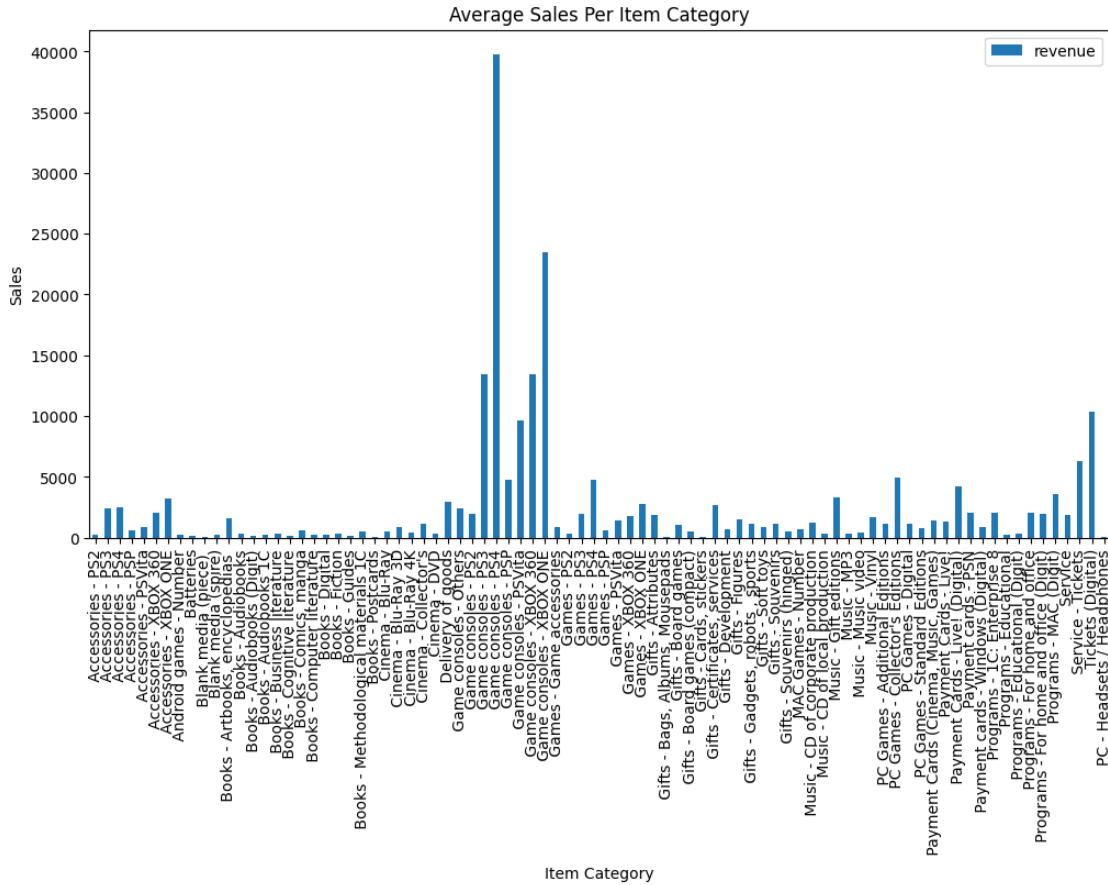
```
[ ]: #Monthly sales Mean, Median, and Standard Deviation
grouped_by_month_name = final_dataset.groupby(['month_name']).agg({'revenue':[
    'mean', 'median', 'std']})
plt.figure(figsize=(12, 6))
sns.lineplot(x=grouped_by_month_name.index,
             y=grouped_by_month_name['revenue']['mean'], label='Mean')
sns.lineplot(x=grouped_by_month_name.index,
             y=grouped_by_month_name['revenue']['median'], label='Median')
sns.lineplot(x=grouped_by_month_name.index,
             y=grouped_by_month_name['revenue']['std'], label='Standard Deviation')
plt.title('Monthly Sales Mean, Median, and Standard Deviation')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.xticks(rotation=45)
plt.legend()
plt.show()
```



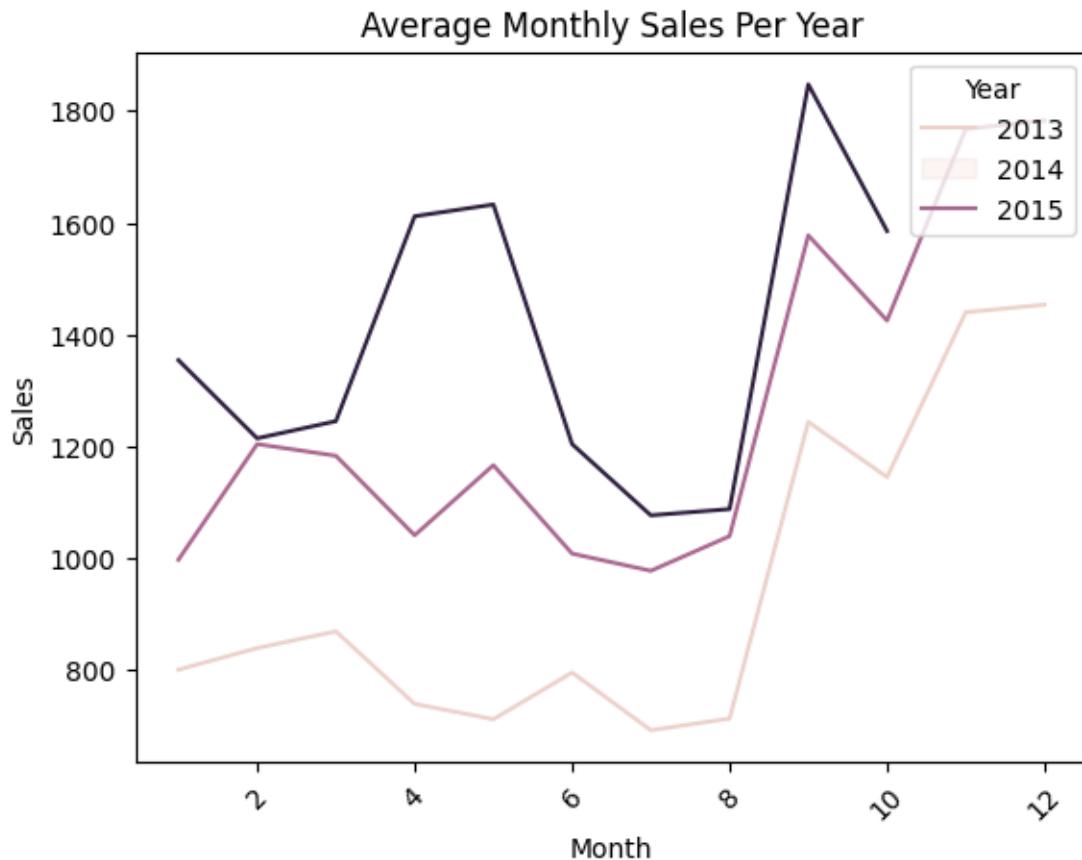
```
[ ]: #popular item categories
grouped_by_item_category_name = final_dataset.groupby(['item_category_name']).  
    agg({'revenue': 'sum'})  
grouped_by_item_category_name.sort_values(by='revenue', ascending=False, □  
    inplace=True)  
plt.figure(figsize=(12, 6))  
plt.pie(grouped_by_item_category_name['revenue'][:10], □  
    labels=grouped_by_item_category_name.index[:10], autopct='%1.1f%%')  
plt.title('Popular Item Categories')  
plt.show()
```



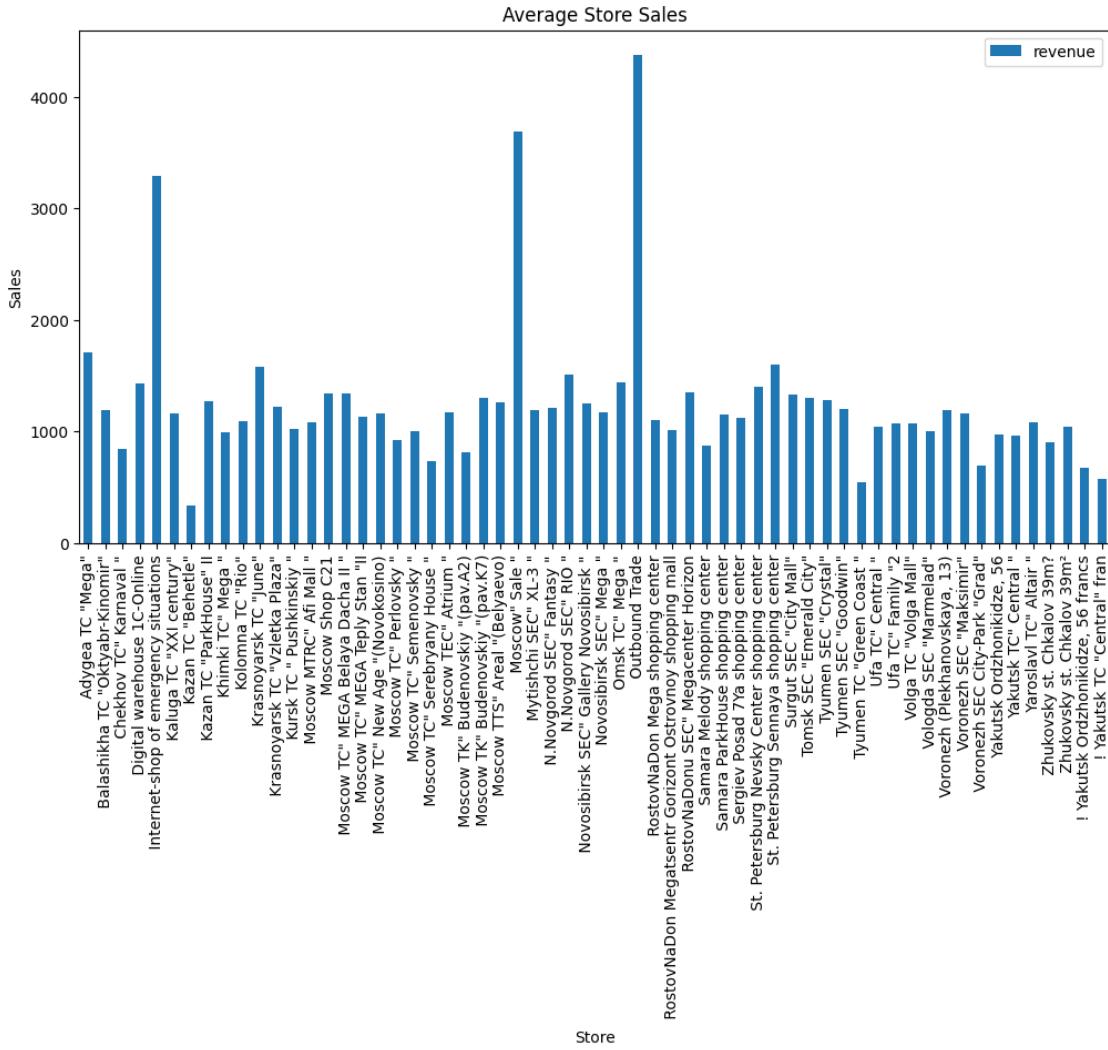
```
[ ]: #average sales per item category
grouped_by_item_category_name = final_dataset.groupby(['item_category_name']).  
    agg({'revenue': 'mean'})  
grouped_by_item_category_name.plot(kind='bar', figsize=(12, 6))  
plt.title('Average Sales Per Item Category')  
plt.xlabel('Item Category')  
plt.ylabel('Sales')  
plt.xticks(rotation=90)  
plt.show()
```



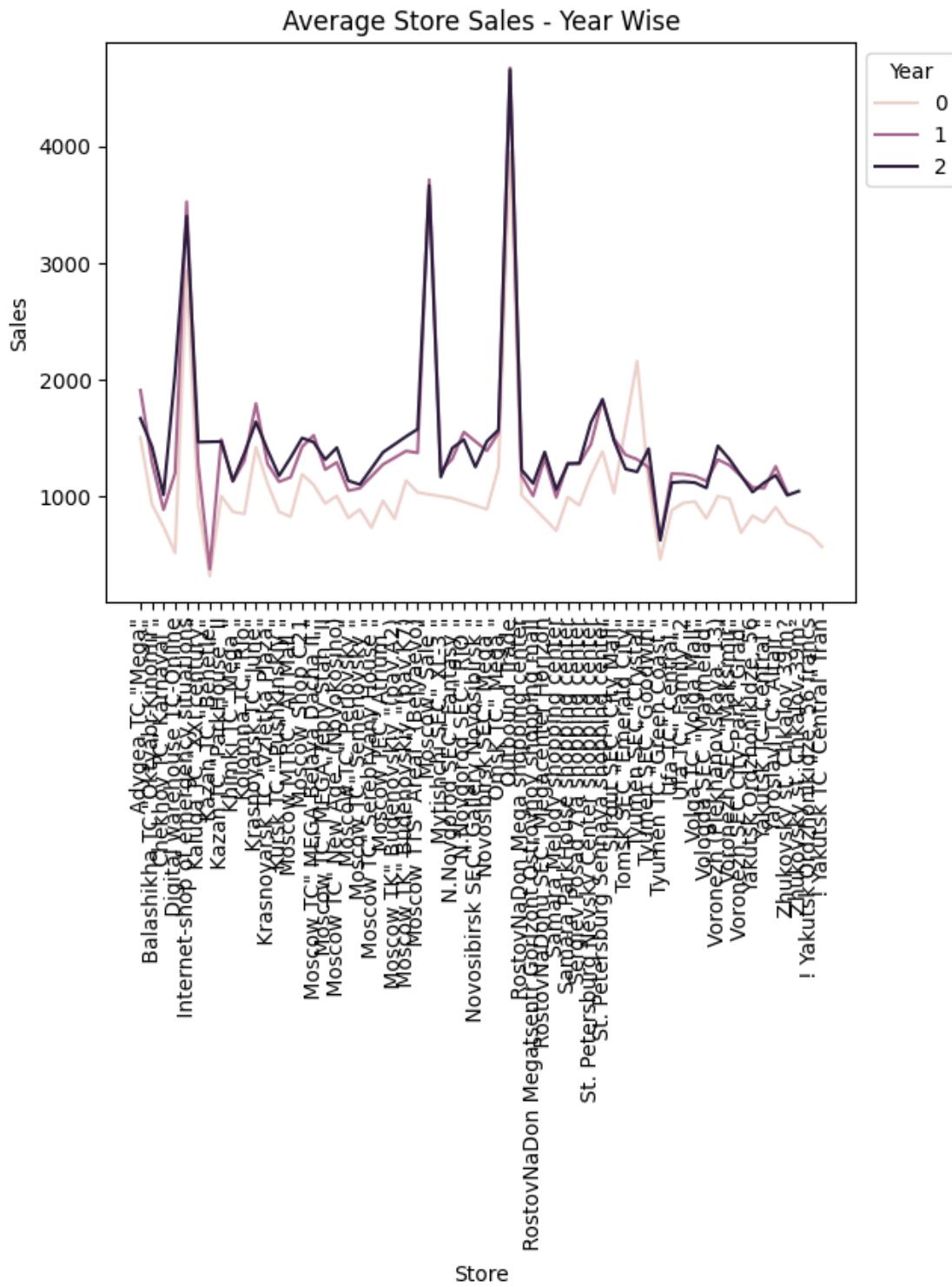
```
[ ]: #average monthly sales per year
grouped_by_year_num_and_month_name = final_dataset.groupby(['year_num', 'month_name']).agg({'revenue': 'mean'})
sns.lineplot(data=grouped_by_year_num_and_month_name, x='month_name', x2='month_name', y='revenue', hue='year_num')
plt.title('Average Monthly Sales Per Year')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.xticks(rotation=45)
plt.legend(title='Year', loc='upper right', labels=['2013', '2014', '2015'])
plt.show()
```



```
[ ]: #average store sales
grouped_by_shop_name = final_dataset.groupby(['shop_name']).agg({'revenue': 'mean'})
grouped_by_shop_name.plot(kind='bar', figsize=(12, 6))
plt.title('Average Store Sales')
plt.xlabel('Store')
plt.ylabel('Sales')
plt.xticks(rotation=90)
plt.show()
```



```
[ ]: #average store sales - year wise
grouped_by_shop_name_and_year_num = final_dataset.groupby(['shop_name', 'year_num']).agg({'revenue': 'mean'})
sns.lineplot(data=grouped_by_shop_name_and_year_num, markers=True, dashes=False, x='shop_name', y='revenue', hue='year_num')
plt.title('Average Store Sales - Year Wise')
plt.xlabel('Store')
plt.ylabel('Sales')
plt.legend(title='Year', loc='upper left', bbox_to_anchor=(1, 1))
plt.xticks(rotation=90)
plt.show()
```

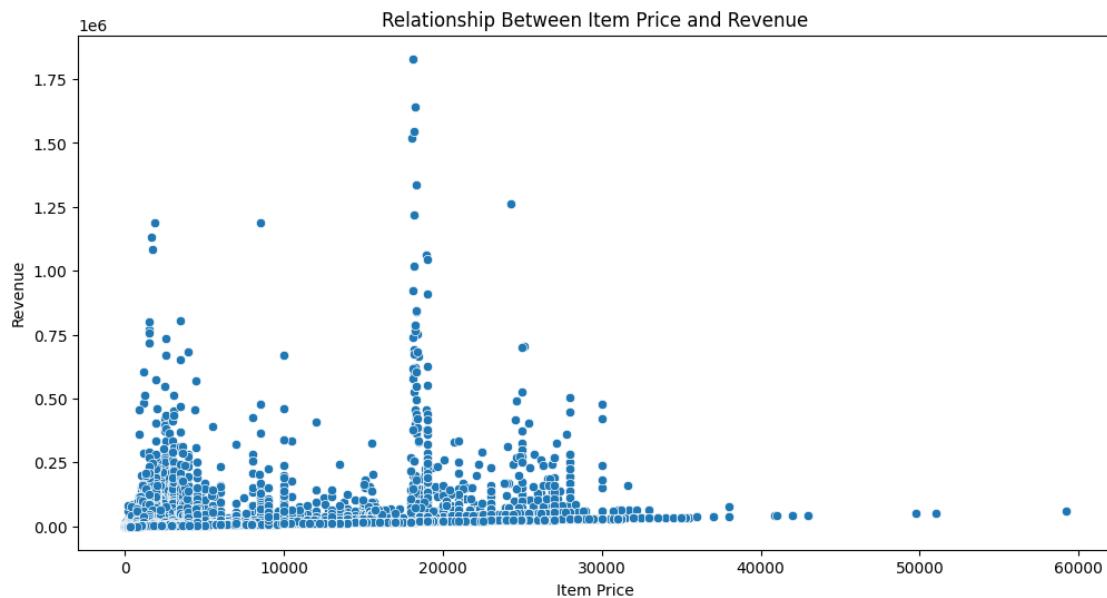


```
[ ]: #relationship between item price and revenue  
plt.figure(figsize=(12, 6))
```

```

sns.scatterplot(data=final_dataset, x='item_price', y='revenue')
plt.title('Relationship Between Item Price and Revenue')
plt.xlabel('Item Price')
plt.ylabel('Revenue')
plt.show()

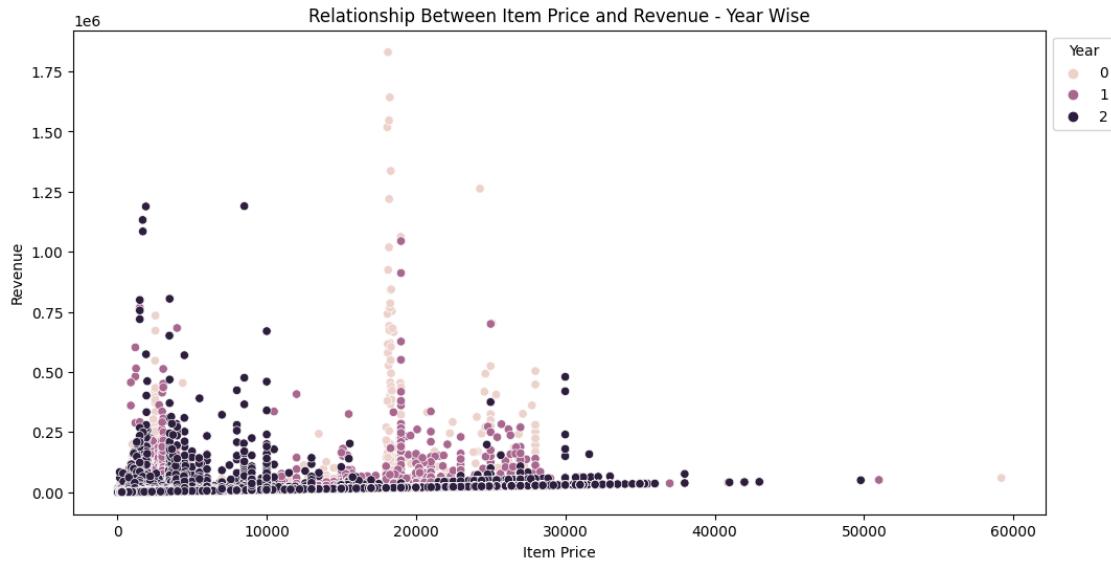
```



```

[ ]: #relationship between item price and revenue - year wise
plt.figure(figsize=(12, 6))
sns.scatterplot(data=final_dataset, x='item_price', y='revenue', hue='year_num')
plt.title('Relationship Between Item Price and Revenue - Year Wise')
plt.xlabel('Item Price')
plt.ylabel('Revenue')
plt.legend(title='Year', loc='upper left', bbox_to_anchor=(1, 1))
plt.show()

```



```
[ ]: #relationship: month of year vs sales
grouped_by_month_name = final_dataset.groupby(['month_name']).agg({'revenue': sum})
sns.lineplot(x=grouped_by_month_name.index, y=grouped_by_month_name['revenue'])
plt.title('Month of Year vs Sales')
plt.xlabel('Month')
plt.ylabel('Sales')
plt.xticks(rotation=45)
plt.show()
```

