import os  
for dirname, \_, filenames in os.walk('/kaggle/input'):  
 for filename in filenames:  
 print(os.path.join(dirname, filename))

import numpy as np  
import pandas as pd  
import seaborn as sns  
import matplotlib.pyplot as plt  
  
color\_pal = sns.color\_palette()  
plt.style.use('fivethirtyeight')

df = pd.read\_csv('/content/drive/MyDrive/Historical Product Demand.csv')

df.head()

Product\_Code Warehouse Product\_Category Date Order\_Demand  
0 Product\_0993 Whse\_J Category\_028 2012/7/27 100   
1 Product\_0979 Whse\_J Category\_028 2012/1/19 500   
2 Product\_0979 Whse\_J Category\_028 2012/2/3 500   
3 Product\_0979 Whse\_J Category\_028 2012/2/9 500   
4 Product\_0979 Whse\_J Category\_028 2012/3/2 500

df.shape

(1048575, 5)

df.columns

Index(['Product\_Code', 'Warehouse', 'Product\_Category', 'Date',  
 'Order\_Demand'],  
 dtype='object')

df.Product\_Code.unique()

array(['Product\_0993', 'Product\_0979', 'Product\_1159', ...,  
 'Product\_0237', 'Product\_0644', 'Product\_0853'], dtype=object)

df.Warehouse.unique()

array(['Whse\_J', 'Whse\_S', 'Whse\_C', 'Whse\_A'], dtype=object)

df.Product\_Category.nunique()

33

df.dtypes

Product\_Code object  
Warehouse object  
Product\_Category object  
Date object  
Order\_Demand object  
dtype: object

def check\_order\_demand(x):  
 try:  
 int(x)  
 except:  
 return False  
 return True  
#Check where Order\_demand is not an integer  
df[~df.Order\_Demand.apply(lambda x: check\_order\_demand(x))].head(6)

Product\_Code Warehouse Product\_Category Date Order\_Demand  
112290 Product\_2169 Whse\_A Category\_024 2012/8/9 (1)  
112307 Product\_2132 Whse\_A Category\_009 2012/11/1 (24)  
112308 Product\_2144 Whse\_A Category\_009 2012/11/1 (24)  
112356 Product\_2118 Whse\_A Category\_009 2012/3/7 (50)  
112357 Product\_2120 Whse\_A Category\_009 2012/3/7 (100)  
112360 Product\_1794 Whse\_A Category\_024 2012/6/28 (1)

def change\_to\_int(x):  
 try:  
 return int(x)  
 except:  
 return int(x[1:-1])  
check = '(10)'  
change\_to\_int(check)

10

df.Order\_Demand = df.Order\_Demand.apply(lambda x: change\_to\_int(x) )

df.describe()

Order\_Demand  
count 1.048575e+06  
mean 4.906977e+03  
std 2.892678e+04  
min 0.000000e+00  
25% 2.000000e+01  
50% 3.000000e+02  
75% 2.000000e+03  
max 4.000000e+06

df = df.rename(columns = {'Product\_Code': 'Code',  
 'Product\_Category':'Category',  
 'Order\_Demand':'Demand'})  
df.head()

Code Warehouse Category Date Demand  
0 Product\_0993 Whse\_J Category\_028 2012/7/27 100  
1 Product\_0979 Whse\_J Category\_028 2012/1/19 500  
2 Product\_0979 Whse\_J Category\_028 2012/2/3 500  
3 Product\_0979 Whse\_J Category\_028 2012/2/9 500  
4 Product\_0979 Whse\_J Category\_028 2012/3/2 500

100 \* df.isna().sum()[3]/ df.shape[0]

1.0718355863910545

df = df.dropna()  
df.isna().sum()

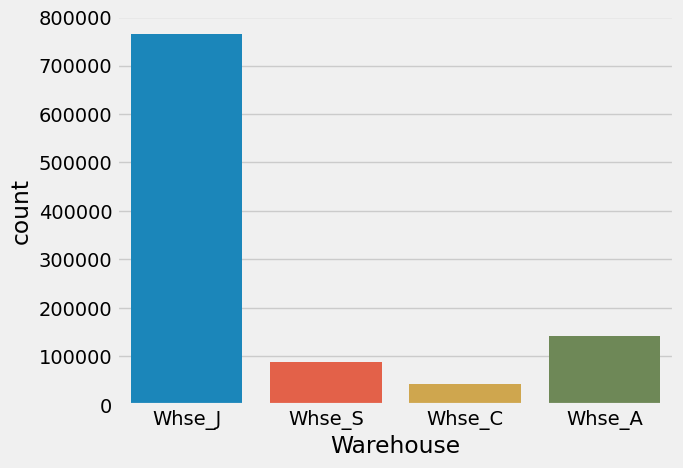
Code 0  
Warehouse 0  
Category 0  
Date 0  
Demand 0  
dtype: int64

df.Date.min(), df.Date.max()

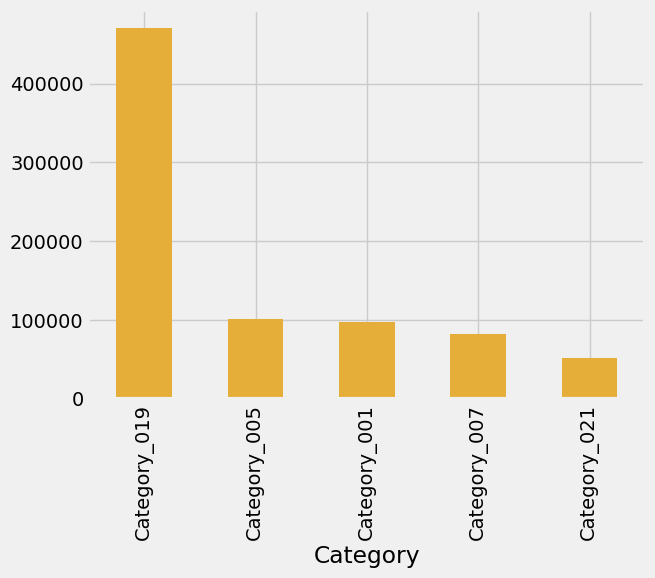
('2011/1/8', '2017/1/9')

# Find the aggregate demand among the warehouses  
sns.countplot(x = 'Warehouse', data = df)

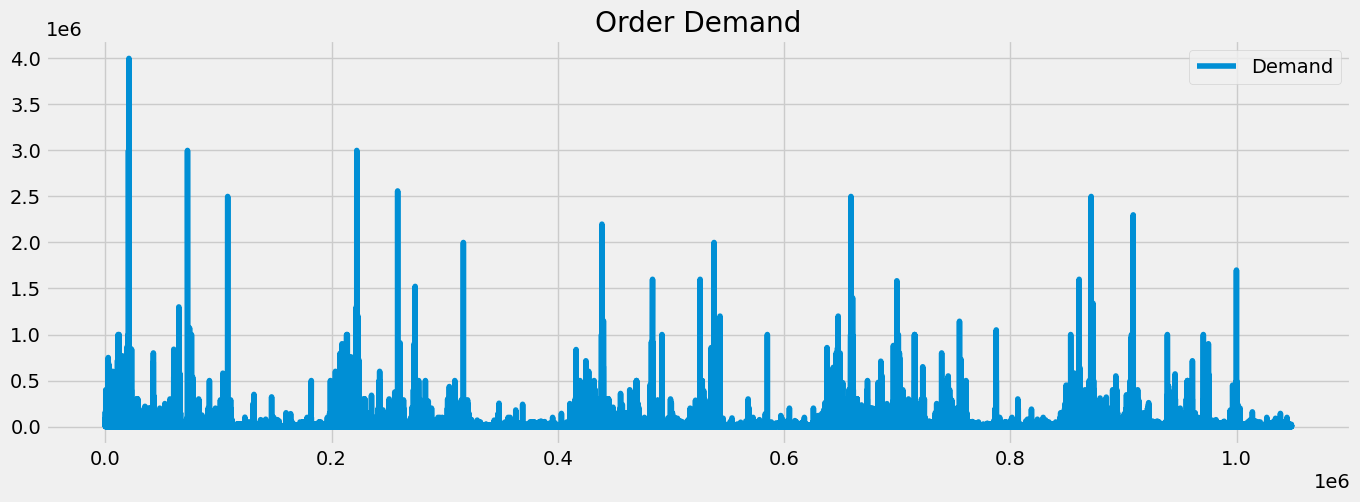
<Axes: xlabel='Warehouse', ylabel='count'>



# Plot the 5 most popular category  
df.Category.value\_counts().head(5).plot(kind = 'bar', color = color\_pal[2])  
plt.xlabel('Category')  
plt.show()



df.plot(kind = 'line',figsize=(15, 5),color = color\_pal[0], title = 'Order Demand' )  
plt.show()



df.Demand.skew()

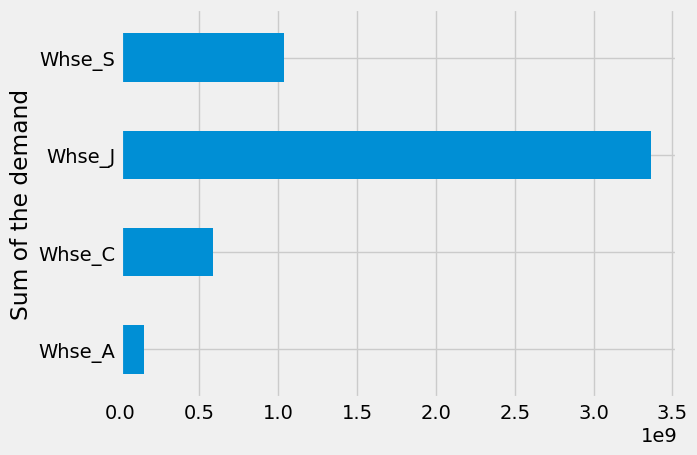
31.432925049321977

# Total Demand by Warehouse  
warehouse\_Demand = df.groupby('Warehouse')['Demand'].sum()  
warehouse\_Demand

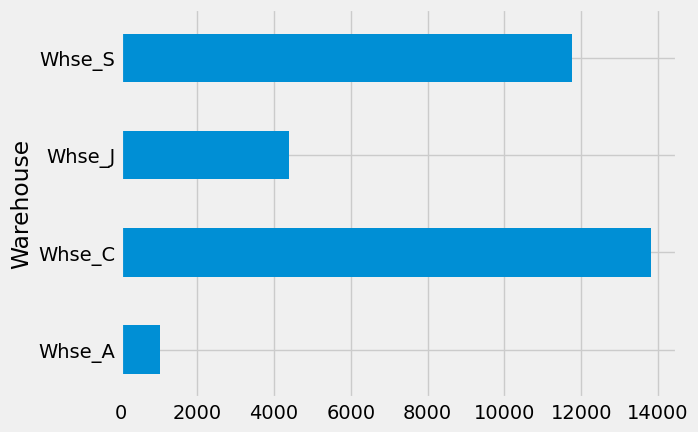
Warehouse  
Whse\_A 147877431  
Whse\_C 585071404  
Whse\_J 3363200396  
Whse\_S 1038024700  
Name: Demand, dtype: int64

warehouse\_Demand.plot(kind = 'barh', ylabel = 'Sum of the demand' )

<Axes: ylabel='Sum of the demand'>



df.groupby('Warehouse')['Demand'].mean().plot(kind = 'barh')  
plt.show()



df.head()

Code Warehouse Category Date Demand  
0 Product\_0993 Whse\_J Category\_028 2012/7/27 100  
1 Product\_0979 Whse\_J Category\_028 2012/1/19 500  
2 Product\_0979 Whse\_J Category\_028 2012/2/3 500  
3 Product\_0979 Whse\_J Category\_028 2012/2/9 500  
4 Product\_0979 Whse\_J Category\_028 2012/3/2 500

# features, Target variable  
Features = ['day\_of\_the\_week', 'Quarter','Month', 'Year', 'Week']  
target = ['Demand']