**NAAN MUDHALVAN**

**DATA ANALYTICS**

**PHASE-4**

**PROJECT-1**

**PROJECT TITLE:PRODUCT SALES ANALYSIS**

**DOMAIN:DATA ANALYTICS**

**NAME:VINOTH.R**

**AGENDA:**

**INTRODUCTION**

**PROJECT OVERVIEW**

**PROJECT DEVELOPMENT**

**PROJECT TASK**

**PROJECT TASK**

**COCLUSION**

**THANKYOU**

**INTRODUCTION:**

**In the realm of data analytics, the development and dataset components of our product sales analysis project represent the foundational pillars upon which our quest for actionable insights and informed decision-making is built. The development aspect encapsulates the strategic orchestration of resources, technologies, and processes, from the provisioning of a data warehousing solution to the integration of advanced analytics tools. Simultaneously, the dataset forms the lifeblood of our analysis, encapsulating a rich repository of transactional and contextual data points, ready to be mined and transformed into meaningful patterns and trends. Through the harmonious synergy of these elements, our project seeks to unlock the full potential of sales data, providing invaluable intelligence that empowers us to optimize product strategies, enhance customer experiences, and drive business growth.**

**PROJECT OVERVIEW:**

**The project aims to develop a comprehensive product sales analysis system utilizing data analytics. The development component involves establishing a data warehousing solution on IBM Cloud with Db2 Warehouse, designing the necessary data schema, setting up ETL processes, ensuring data security, and integrating with relevant analytics tools for effective reporting and visualization. Simultaneously, the project will curate and manage a robust dataset that encompasses key sales metrics such as product IDs, unit prices, quantities sold, customer information, payment methods, and location data, ensuring data quality and reliability. This integrated approach will enable insightful and actionable sales analysis, fostering data-driven decision-making within the organization.**

**PROJECT DEVELOPMENT:**

**the development and dataset components of our product sales analysis project represent the foundational pillars upon which our quest for actionable insights and informed decision-making is built. The development aspect encapsulates the strategic orchestration of resources, technologies, and processes, from the provisioning of a data warehousing solution to the integration of advanced analytics tools. Simultaneously, the dataset forms the lifeblood of our analysis, encapsulating a rich repository of transactional and contextual data points, ready to be mined and transformed into meaningful patterns and trends. Through the harmonious synergy of these elements, our project seeks to unlock the full potential of sales data, providing invaluable intelligence that empowers us to optimize product strategies, enhance customer experiences, and drive business growth.**

**DATA SET:**

Customers Demographic Analysis

We will analyze sales by demographic Analysis of customers eg city, age, gender .. The goal of this process is to give more information about our data so that the marketing team prepares to intensify the efficiency based on the data and information we will provide !  
At the end of this intention there is a challenge or duty you have to complete the end you have to understand all the facts that have already occurred and all the facts that I did not mention some and listed as a story you have to tell every part even if simple Ithakk to make a full page as a full report on this subject, From this core is teaching you how exploratory analysis is how you tell stories how to make information in your mind

* [Gender](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#gender)
* [Age](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#age)
* [City](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#city_category)
* [Stability](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#stability)
* [Occupation](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#occupation)
* [Products](https://www.kaggle.com/code/loaiabdalslam/sales-data-analysis-report#products)

In [1]:

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt *# visualizing data*

import seaborn as sns

from collections import Counter

%matplotlib inline

import plotly.plotly as py

from plotly.offline import init\_notebook\_mode, iplot

import plotly.graph\_objs as go

import plotly.figure\_factory as ff

import os

print(os.listdir("../input"))

import plotly.plotly as py

import plotly.graph\_objs as go

import seaborn as sns

['BlackFriday.csv']

In [2]:

df = pd.read\_csv('../input/BlackFriday.csv')

df.shape

Out[2]:

(537577, 12)

In [3]:

*#df.info()*

In [4]:

df.describe()

Out[4]:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 3 | 1000001 | P00085442 | F | 0-17 | 10 | A | 2 | 0 | 12 | 14.0 | NaN | 1057 |
| 4 | 1000002 | P00285442 | M | 55+ | 16 | C | 4+ | 0 | 8 | NaN | NaN | 7969 |

1- Gender

In [6]:

explode = (0.1,0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df['Gender'].value\_counts(), explode=explode,labels=['Male','Female'], autopct='**%1.1f%%**',

shadow=True, startangle=90)

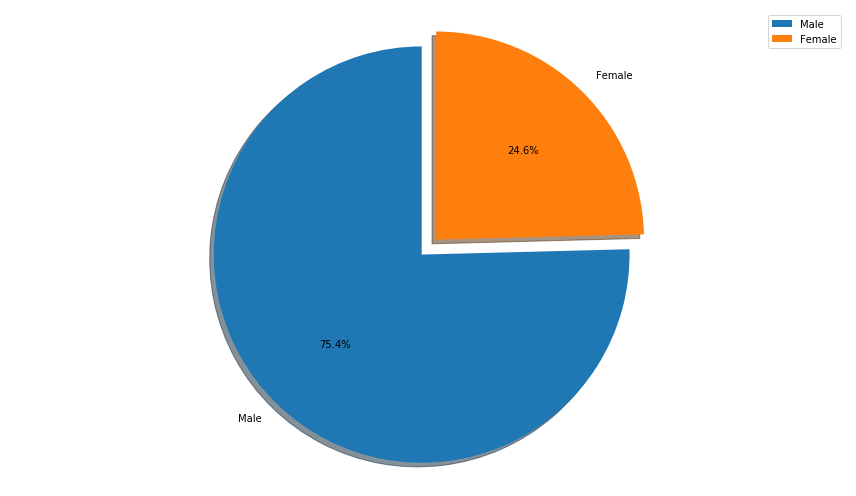
*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

plt.legend()

plt.show()



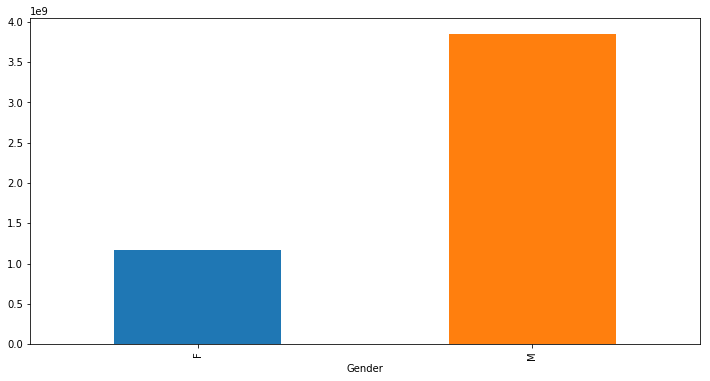
In [7]:

def plot(group,column,plot):

ax=plt.figure(figsize=(12,6))

df.groupby(group)[column].sum().sort\_values().plot(plot)

plot('Gender','Purchase','bar')



**Men's purchasing power is greater than women's purchasing power, even in normal circumstances. This is likely to affect the owner of the money, but there has been a high turnout of men in the store. About 75% of the customers have made sales of men of all ages, The men are generally heading toward products at 8000 - 12,000, we have probably made sales worth more than 4 billion in men and more than 1 billion in ladies**

2-Age

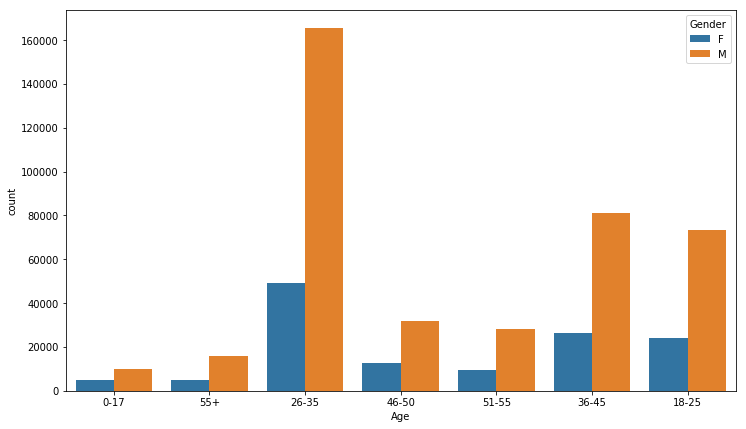
In [8]:

fig1, ax1 = plt.subplots(figsize=(12,7))

sns.countplot(df['Age'],hue=df['Gender'])

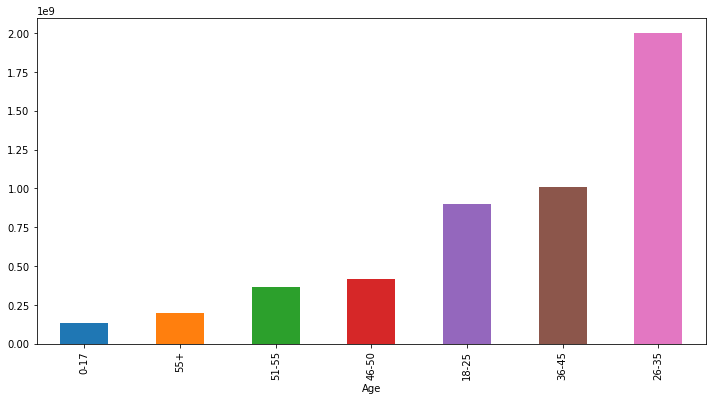
Out[8]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fba4dfa0a58>



In [9]:

plot('Age','Purchase','bar')



**Obviously, we can consider that the target age group of our stores is the age group of 26-35 years, we have achieved sales of more than 3 billion in the age group of 26-45 years**

3-City

In [10]:

explode = (0.1, 0, 0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df['City\_Category'].value\_counts(),explode=explode, labels=df['City\_Category'].unique(), autopct='**%1.1f%%**',

shadow=True, startangle=90)

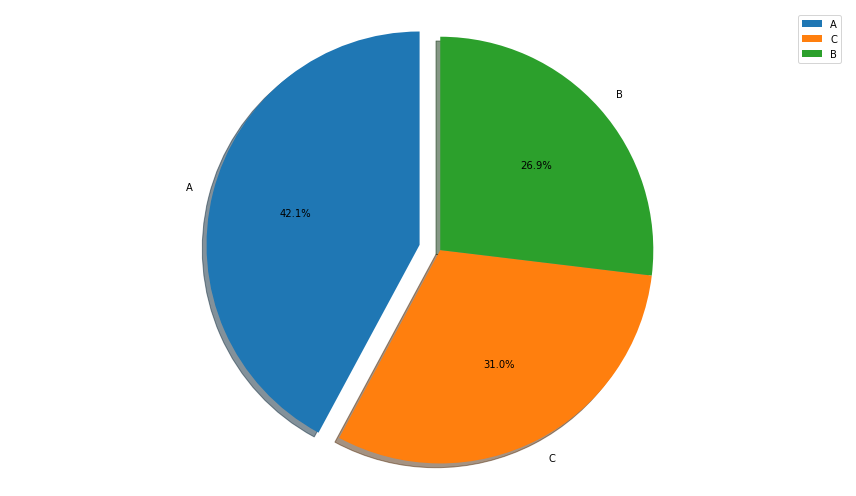
*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

plt.legend()

plt.show()



In [11]:

explode = (0.1, 0, 0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df.groupby('City\_Category')['Purchase'].sum(),explode=explode, labels=df['City\_Category'].unique(), autopct='**%1.1f%%**',

shadow=True, startangle=90)

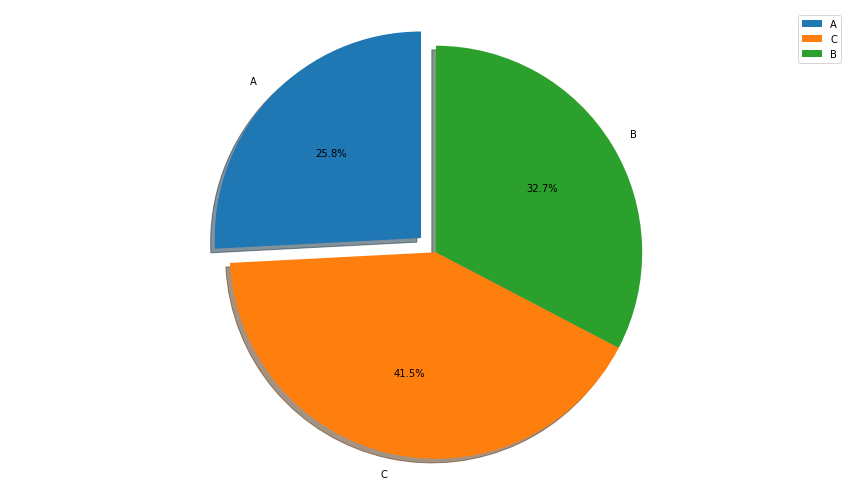
*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

plt.legend()

plt.show()



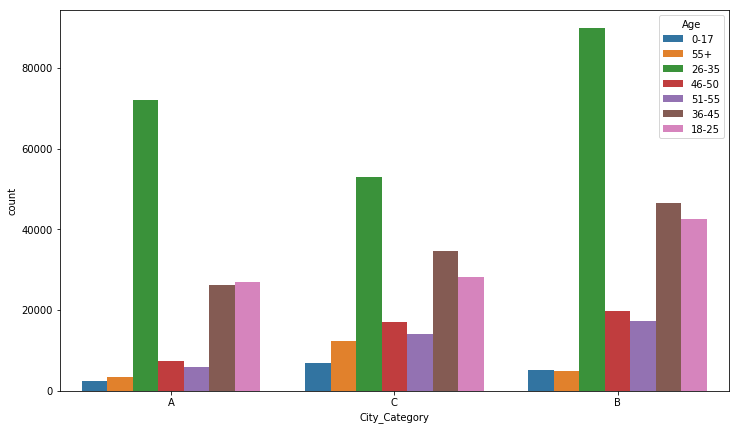
In [12]:

fig1, ax1 = plt.subplots(figsize=(12,7))

sns.countplot(df['City\_Category'],hue=df['Age'])

Out[12]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fba4de582e8>



**Unexpectedly, the highest sales do not come in the number of purchases, people from Area B have a greater purchasing power than others, and greater sales gained from people from Area C**

Look at the comments, sometimes we have to identify the appropriate wording and not mix with it as I did here. I would like to point out that sales do not reflect purchasing power, but the number of attendees reflects purchasing power because the data are individual sales. sales in the city, it reflects the purchasing power ..

In [13]:

*#label=['Underage 0-17','Retired +55','Middleage 26-35','46-50 y/o','Oldman 51-55','Middleage+ 36-45','Youth']*

explode = (0.1, 0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df['Marital\_Status'].value\_counts(),explode=explode, labels=['Yes','No'], autopct='**%1.1f%%**',

shadow=True, startangle=90)

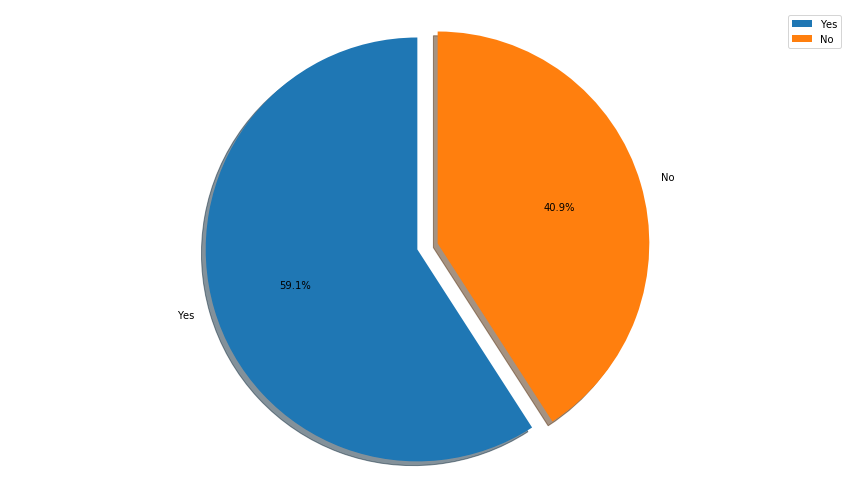
*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

plt.legend()

plt.show()



**City A is the most cities followed by B and then C, the distribution of ages on the procurement map is very close, we have to focus on that category of work averages of 36-45**

**Most of our customers are more than 60% married, I see that the strategy of targeting families to ensure more clients succeed**

4-Stability

In [14]:

labels=['First Year','Second Year','Third Year','More Than Four Years','Geust']

explode = (0.1, 0.1,0,0,0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df.groupby('Stay\_In\_Current\_City\_Years')['Purchase'].sum(),explode=explode, labels=labels, autopct='**%1.1f%%**',

shadow=True, startangle=90)

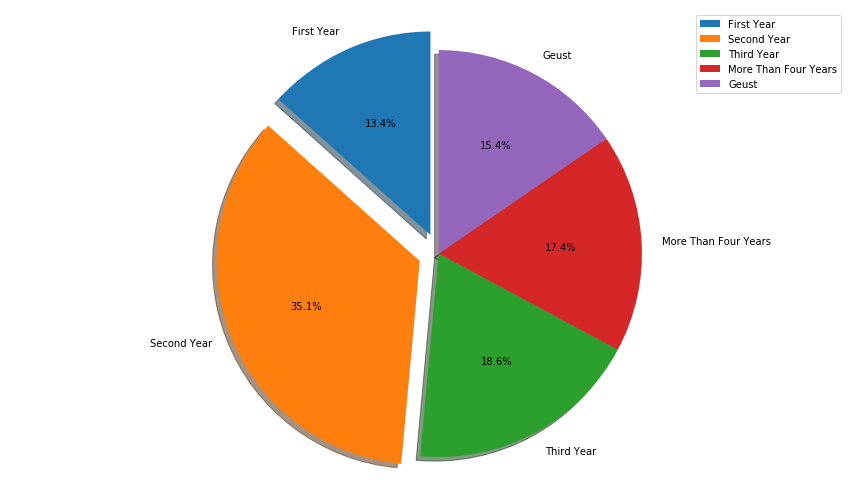
*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

plt.legend()

plt.show()



In [15]:

labels=['First Year','Second Year','Third Year','More Than Four Years','Geust']

*#label=['Underage 0-17','Retired +55','Middleage 26-35','46-50 y/o','Oldman 51-55','Middleage+ 36-45','Youth']*

explode = (0.1, 0.1,0,0,0)

fig1, ax1 = plt.subplots(figsize=(12,7))

ax1.pie(df['Stay\_In\_Current\_City\_Years'].value\_counts(),explode=explode, labels=labels, autopct='**%1.1f%%**',

shadow=True, startangle=90)

*# Equal aspect ratio ensures that pie is drawn as a circle*

ax1.axis('equal')

plt.tight\_layout()

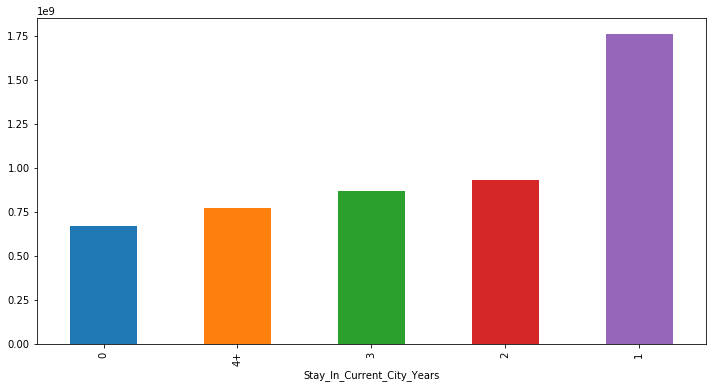
plt.legend()

plt.show()

unfold\_moreShow hidden output

In [16]:

plot('Stay\_In\_Current\_City\_Years','Purchase','bar')



**We have worked hard in the past two years and have achieved a large percentage of sales from the new population of cities, but these figures indicate that the older city dwellers have less passion for our products. I do not know in fact look at it for yourselves why old city dwellers did not achieve higher sales of the population New visitors or visitors from outside the city?  
We have almost gained about 1.75 billion new city residents only!**

5-Occupation

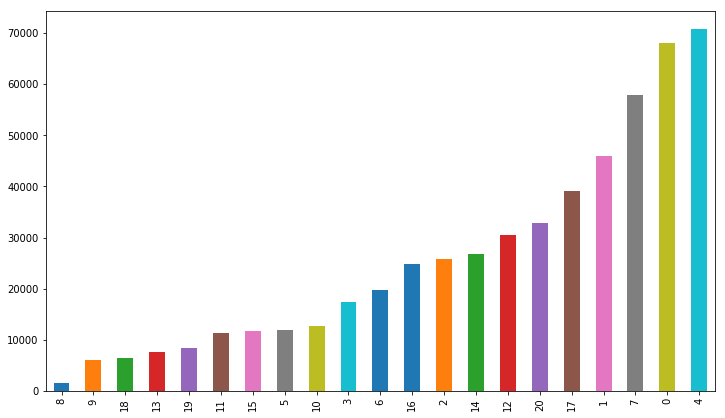
In [17]:

fig1, ax1 = plt.subplots(figsize=(12,7))

df['Occupation'].value\_counts().sort\_values().plot('bar')

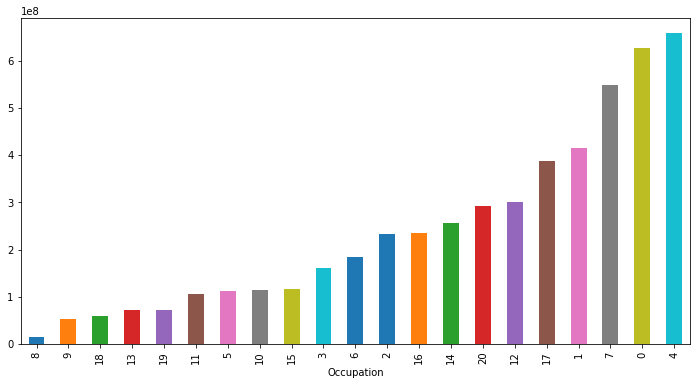
Out[17]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fba4dbe6400>



In [18]:

plot('Occupation','Purchase','bar')

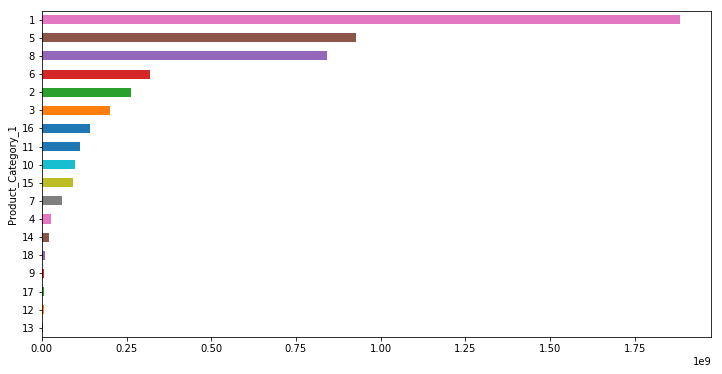


**We also note here that purchasing power is closely related to the Occupation in some cases as the first class of the table but there are some differences we will notice when checking the number of purchases and the value of those purchases total**

6-Products and Catiegories

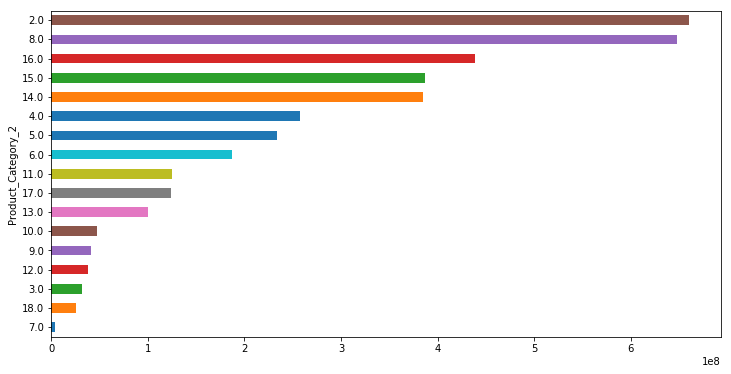
In [19]:

plot('Product\_Category\_1','Purchase','barh')



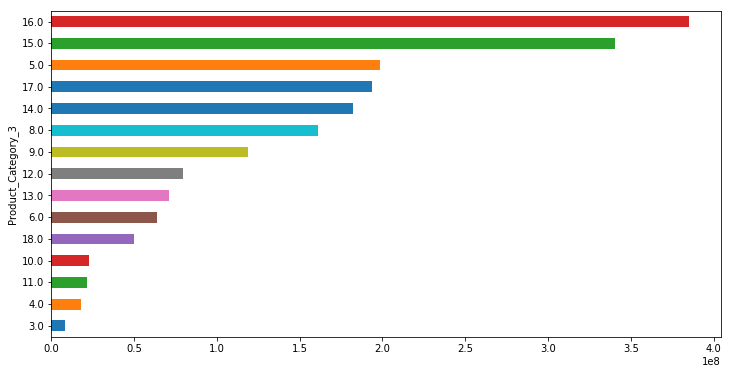
In [20]:

plot('Product\_Category\_2','Purchase','barh')



In [21]:

plot('Product\_Category\_3','Purchase','barh')



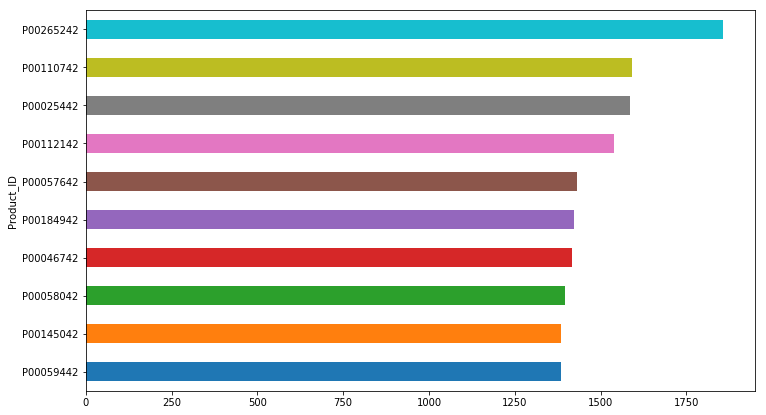
In [22]:

fig1, ax1 = plt.subplots(figsize=(12,7))

df.groupby('Product\_ID')['Purchase'].count().nlargest(10).sort\_values().plot('barh')

Out[22]:

<matplotlib.axes.\_subplots.AxesSubplot at 0x7fba4d9135c0>



In [23]:

fig1, ax1 = plt.subplots(figsize=(12,7))

**CONCLUSION:**

**In conclusion, the development and dataset creation stages of the product sales analysis project using data analytics are foundational pillars of our endeavor. The robustness of our data warehousing infrastructure, established through IBM Cloud Db2 Warehouse, provides a secure and scalable environment for data storage and retrieval. The meticulously crafted dataset, featuring key attributes such as product details, sales transactions, and customer information, forms the bedrock for insightful analytics. By meticulously planning, designing, and implementing these essential elements, we set the stage for comprehensive, data-driven insights, enabling us to make informed business decisions, identify trends, and optimize our sales strategies with a keen eye on customer needs and market dynamics.**

**THANKYOU**