FACULTY OF ENGINEERING ANDTECHNOLOGY

OF

TECHNOLOGY

Data Analytics And Data visualization laboratory

Subject Code:303105315

3rd year 5th Sem

Laboratory Manual-DADV

Enrollment No: xxxxxxxxx

CERTIFICATE

This is to certify that **Mr.xxxxxxx** with Enrolment **No.xxxxxx** and 5th Semester/ CSE (Batch No) has successfully completed her laboratory experiments in the **Data Analytics And Data visualization laboratory** (303105315) from the **Department of Computer Science & Engineering duringthe academic year 2024-25.**

Date of Submission: ....

Staff In charge: ....

Head Of Department: ....

TABLE OF CONTENT

Practical-1

Aim: Perform Exploratory Data Analysis on the given dataset using Python.

Procedure:

1.Import the dataset
2.View the head of the data
3.View the basic information of data and description of data
4.Find the unique value of data and verify the duplication of data
5.Plot a graph for unique value of dataset
6.Verify the presence of null value and replace the null value
7.Visualize the needed data

Program:

```
#Load the required libraries
import pandas as pd
import numpy as  np

import seaborn as sns
#Load the data

df  =  pd.read_csv('titanic.csv')


#View the data
df.head()

df.info()
df.describe()
#Find the duplicates
df.duplicated().sum()

#unique values
df['Pclass'].unique()
df['Survived'].unique()
df['Sex'].unique()
          array([3, 1, 2], dtype=int64)
          array([0, 1], dtype=int64)
          array(['male', 'female'], dtype=object)

#Plot the unique values
sns.countplot(df['Pclass']).unique()

#Find null values
df.isnull().sum()

#Replace null values
df.replace(np.nan,'0',inplace = True)
```

```
#Check the changes now
df.isnull().sum()

#Filter data
df[df['Pclass']==1].head()

#Boxplot
df[['Fare']].boxplot()
```

Practical-2

Aim: Calculate mean, median and mode of the first 50 records in the given dataset using python

```
from statistics import median

import pandas as pd

import matplotlib.pyplot as plt

d = pd.read_excel(r'F:\SEMESTER 5\DATA VISULIZATION AND DATA
ANALYTICS\LAB\Marks.x

lsx')f = pd.DataFrame(d)

print(f)

meanm = d["Maths"].mean()

medianm = d["Maths"].median()

modem = d["Maths"].mode()

meanp = d["Physics"].mean()

medianp =

d["Physics"].median()modep =

d["Physics"].mode() cor =

d.corr()

print("Maths:")

print("Mean: ",meanm)

print("Median: ",medianm)

print("Mode: ",modem)

print("Physics:")

print("Mean: ",meanp)

print("Median: ",medianp)

print("Mode: ",modep)
```

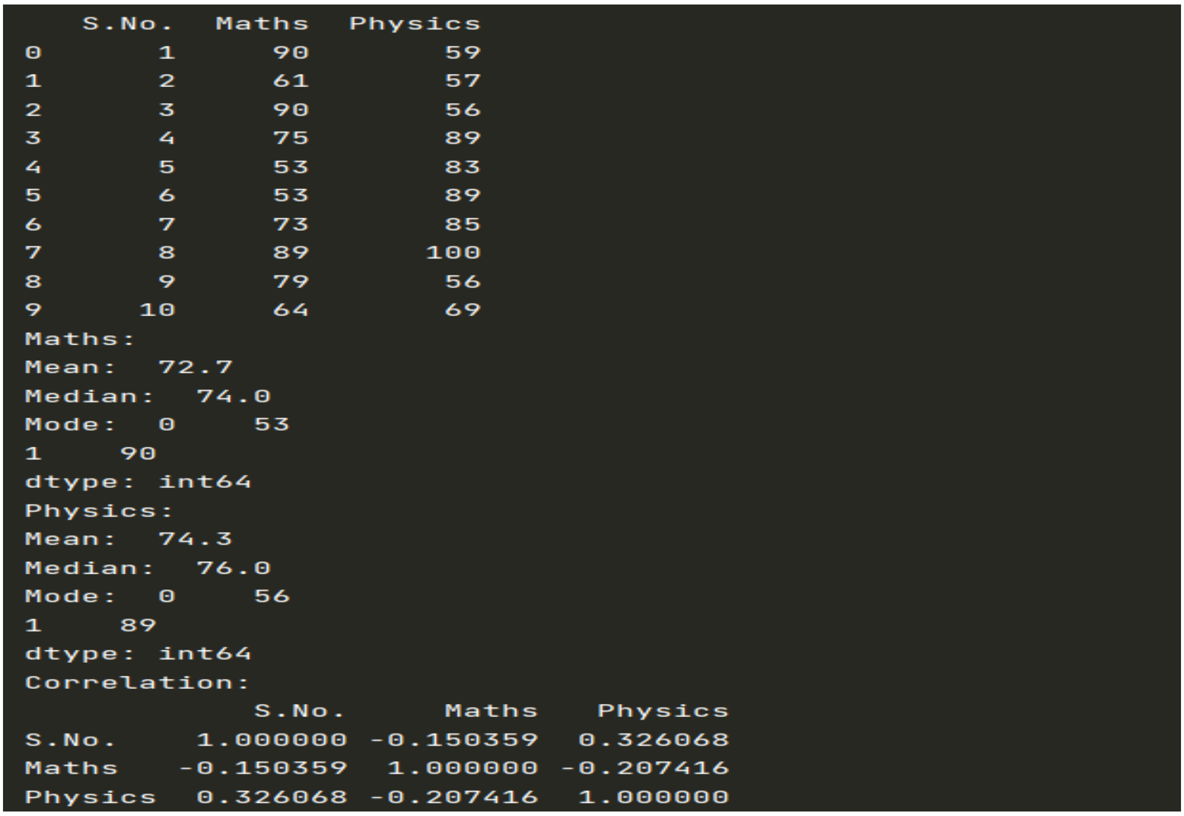
print("Correlation: ")

print(cor)

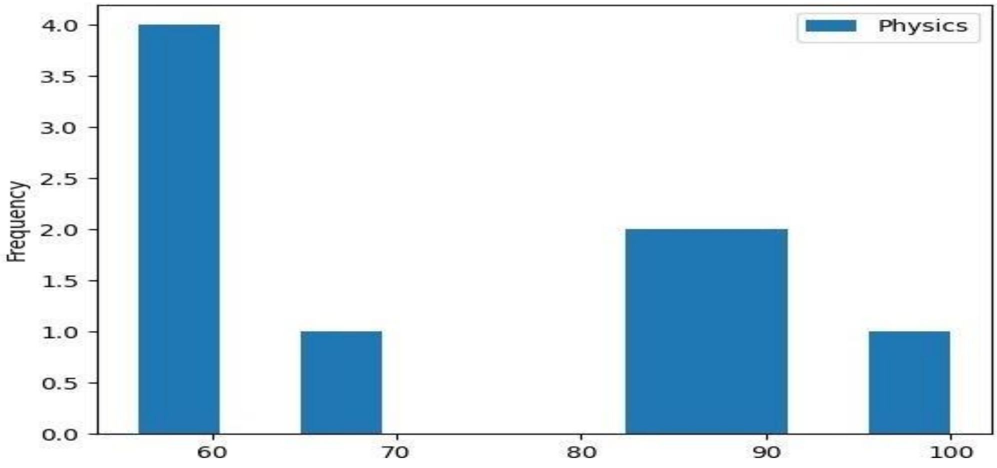
d.plot(kind = 'hist', x = 'Maths', y =

'Physics')plt.show()

OUTPUT:



```
   S.No.  Maths  Physics
0      1     90       59
1      2     61       57
2      3     90       56
3      4     75       89
4      5     53       83
5      6     53       89
6      7     73       85
7      8     89      100
8      9     79       56
9     10     64       69
Maths:
Mean:  72.7
Median:  74.0
Mode:  0    53
1    90
dtype: int64
Physics:
Mean:  74.3
Median:  76.0
Mode:  0    56
1    89
dtype: int64
Correlation:
            S.No.     Maths   Physics
S.No.    1.000000 -0.150359  0.326068
Maths   -0.150359  1.000000 -0.207416
Physics  0.326068 -0.207416  1.000000
```

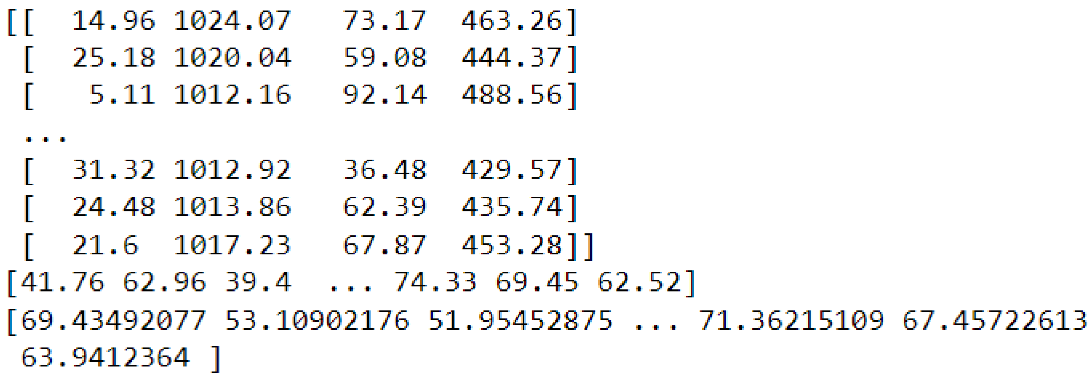


Practical-3

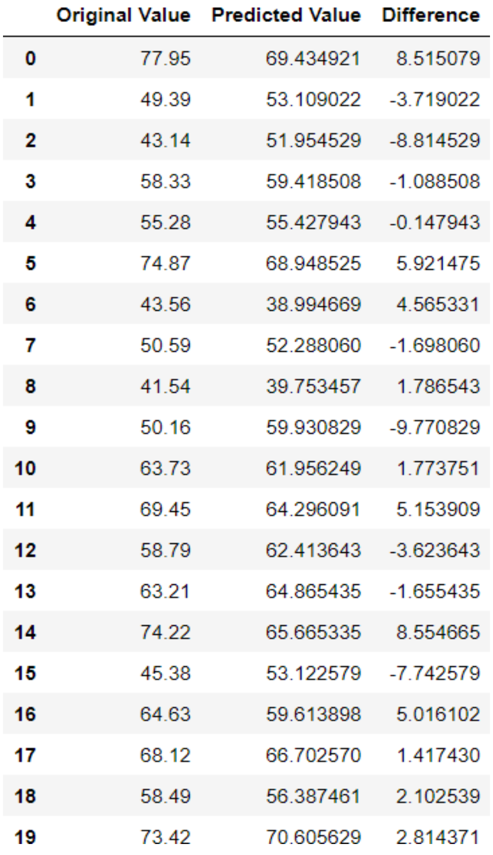
Aim: Perform Multiple Linear Regression on data.

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
import matplotlib.pyplot as plt
df=pd.read_excel(r'H:\CODES\PYTHON\JEET.xlsx')
df.head()
x=df.drop(['V'],axis=1).value
sy=df['V'].values
print(x)
print(y)
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
ml=LinearRegression()
ml.fit(x_train,y_train)
y_pred=ml.predict(x_test)
print(y_pred)
r2_score(y_test,y_pred)
plt.figure(figsize=(15,10))
plt.scatter(y_test,y_pred)
plt.xlabel('Original')
plt.ylabel('Predicted')

plt.title('Original vs Predicted')

pred_y_df=pd.DataFrame({'Original Value':y_test,'Predicted Value':y_pred,'Difference':y_test-y_pred})

pred_y_df[0:20]
```

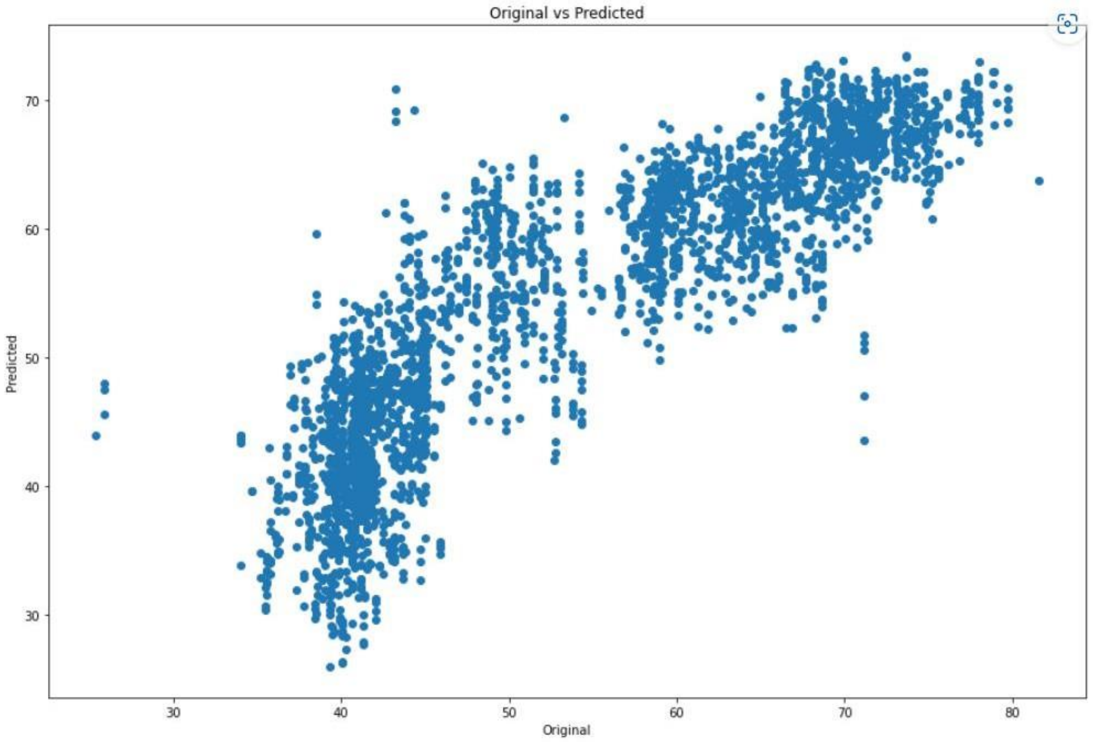
OUTPUT



```
[[   14.96 1024.07    73.17   463.26]
 [   25.18 1020.04    59.08   444.37]
 [    5.11 1012.16    92.14   488.56]
 ...
 [   31.32 1012.92    36.48   429.57]
 [   24.48 1013.86    62.39   435.74]
 [   21.6  1017.23    67.87   453.28]]
[41.76 62.96 39.4  ... 74.33 69.45 62.52]
[69.43492077 53.10902176 51.95452875 ... 71.36215109 67.45722613
 63.9412364 ]
```



	Original Value	Predicted Value	Difference
0	77.95	69.434921	8.515079
1	49.39	53.109022	-3.719022
2	43.14	51.954529	-8.814529
3	58.33	59.418508	-1.088508
4	55.28	55.427943	-0.147943
5	74.87	68.948525	5.921475
6	43.56	38.994669	4.565331
7	50.59	52.288060	-1.698060
8	41.54	39.753457	1.786543
9	50.16	59.930829	-9.770829
10	63.73	61.956249	1.773751
11	69.45	64.296091	5.153909
12	58.79	62.413643	-3.623643
13	63.21	64.865435	-1.655435
14	74.22	65.665335	8.554665
15	45.38	53.122579	-7.742579
16	64.63	59.613898	5.016102
17	68.12	66.702570	1.417430
18	58.49	56.387461	2.102539
19	73.42	70.605629	2.814371


Original vs Predicted
Predicted
Original
30
40
50
60
70
80

Practical-4

Aim: Perform the Logistic Regression on a dataset.

LOGISTIC REGRESSION: Logistic regression estimates the probability of anevent occurring, such as voted or didn't vote, based on a given dataset of independent variables

CODE:

```
from sklearn.datasets import load_digits

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

digits = load_digits()

dir(digits)

digits.data[4]

plt.gray()

plt.matshow(digits.images[1])

digits.target[0:5]

x_train,x_test,y_train,y_test = train_test_split(digits.data,digits.target,test_size=0.2)

len(x_test)

logistic = LogisticRegression()

logistic.fit(x_train,y_train)

logistic.score(x_test,y_test)

plt.matshow(digits.images[84])

digits.target[84]
```

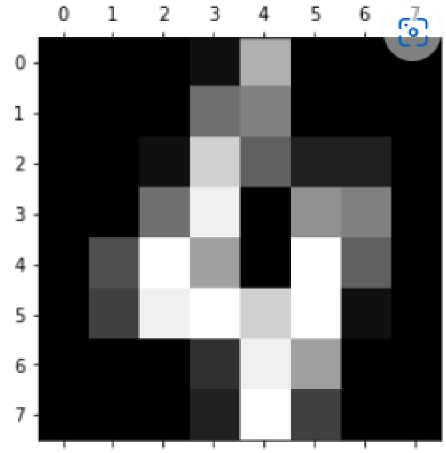
logistic.predict([digits.data[84]])

OUTPUT

```
['DESCR', 'data', 'feature_names', 'frame', 'images', 'target', 'target_names']
```

```
array([ 0.,  0.,  0.,  1., 11.,  0.,  0.,  0.,  0.,  0.,  0.,  7.,  8.,
        0.,  0.,  0.,  0.,  0.,  1., 13.,  6.,  2.,  2.,  0.,  0.,  0.,
        7., 15.,  0.,  9.,  8.,  0.,  0.,  5., 16., 10.,  0., 16.,  6.,
        0.,  0.,  4., 15., 16., 13., 16.,  1.,  0.,  0.,  0.,  0.,  3.,
       15., 10.,  0.,  0.,  0.,  0.,  0.,  2., 16.,  4.,  0.,  0.])
```

```
: <matplotlib.image.AxesImage at 0x24a08d62e50>

  <Figure size 432x288 with 0 Axes>
```

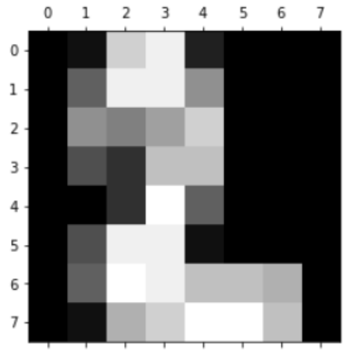


```
array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])
```

```
360
```

```
0.9444444444444444
```

```
<matplotlib.image.AxesImage at 0x24a08b2b5b0>
```



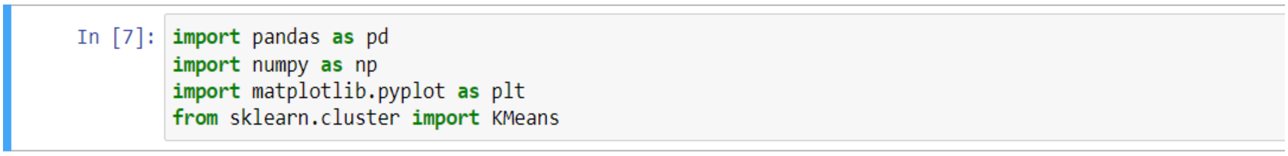
```
array([2])
```

Practical-5

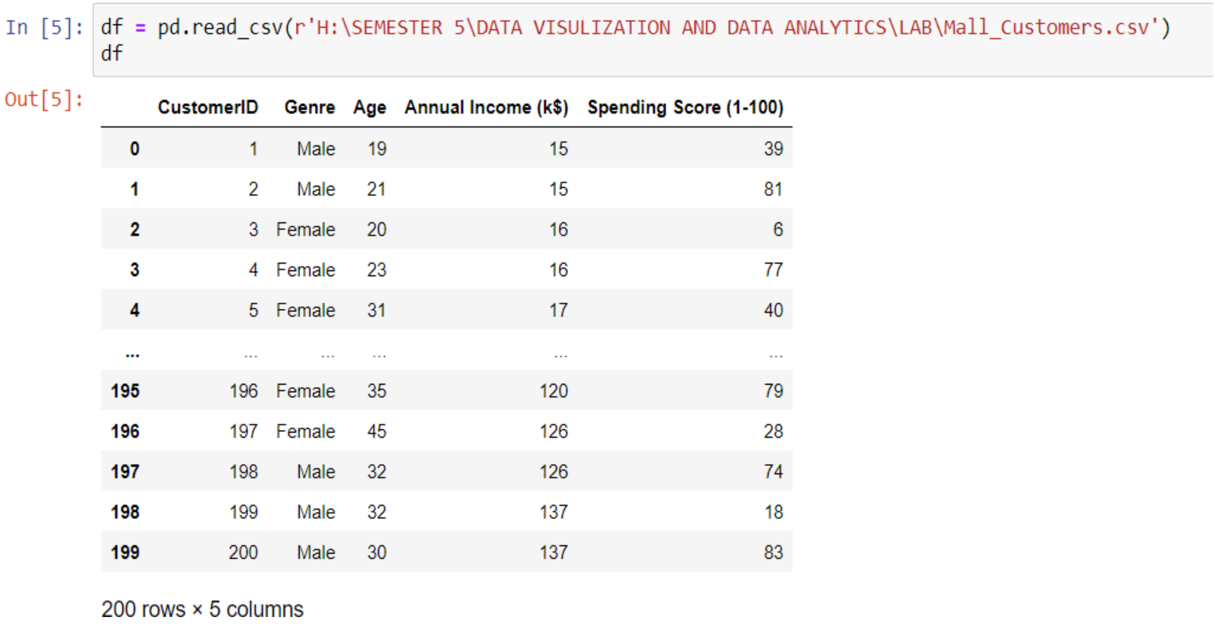
Aim: Use a dataset & apply K means clustering to get insights from data.

K MEANS CLUSTING: K means clustering is one of the simplest algorithm whichuses unsupervised learning method to solve known clustering issues

CODE:



```
In [7]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        from sklearn.cluster import KMeans
```

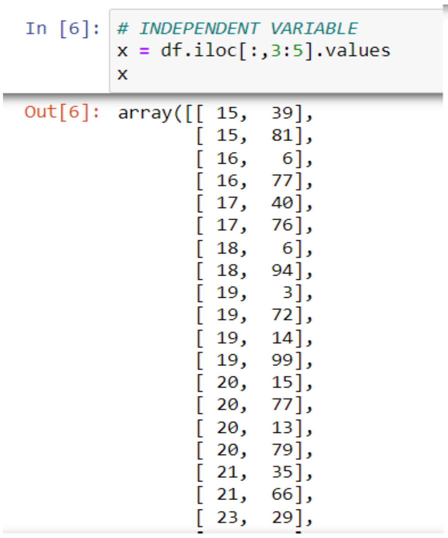


```
In [5]: df = pd.read_csv(r'H:\SEMESTER 5\DATA VISULIZATION AND DATA ANALYTICS\LAB\Mall_Customers.csv')
        df
```

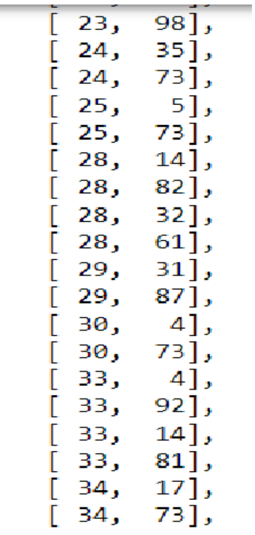
Out[5]:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

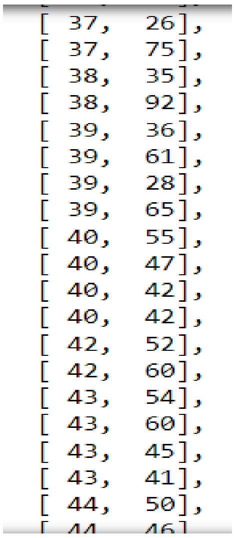
200 rows × 5 columns



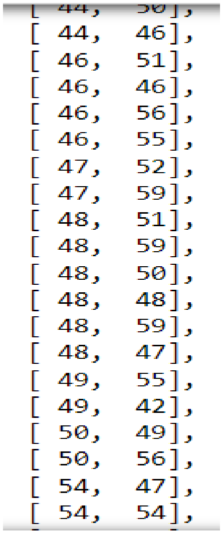
```
In [6]: # INDEPENDENT VARIABLE
        x = df.iloc[:,3:5].values
        x

Out[6]: array([[ 15,  39],
               [ 15,  81],
               [ 16,   6],
               [ 16,  77],
               [ 17,  40],
               [ 17,  76],
               [ 18,   6],
               [ 18,  94],
               [ 19,   3],
               [ 19,  72],
               [ 19,  14],
               [ 19,  99],
               [ 20,  15],
               [ 20,  77],
               [ 20,  13],
               [ 20,  79],
               [ 21,  35],
               [ 21,  66],
               [ 23,  29],
```



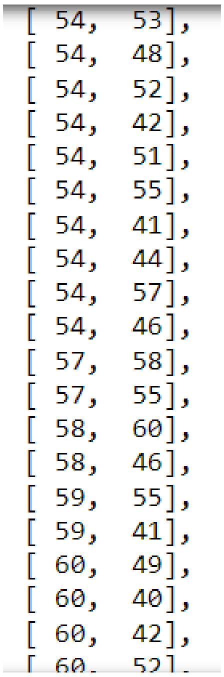
```
[ 23,   98],
[ 24,   35],
[ 24,   73],
[ 25,    5],
[ 25,   73],
[ 28,   14],
[ 28,   82],
[ 28,   32],
[ 28,   61],
[ 29,   31],
[ 29,   87],
[ 30,    4],
[ 30,   73],
[ 33,    4],
[ 33,   92],
[ 33,   14],
[ 33,   81],
[ 34,   17],
[ 34,   73],
```



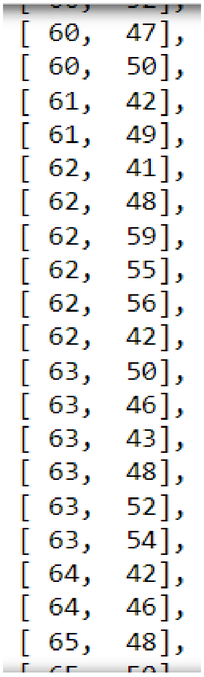
```
[ 37,   26],
[ 37,   75],
[ 38,   35],
[ 38,   92],
[ 39,   36],
[ 39,   61],
[ 39,   28],
[ 39,   65],
[ 40,   55],
[ 40,   47],
[ 40,   42],
[ 40,   42],
[ 42,   52],
[ 42,   60],
[ 43,   54],
[ 43,   60],
[ 43,   45],
[ 43,   41],
[ 44,   50],
[ 44,   46]
```



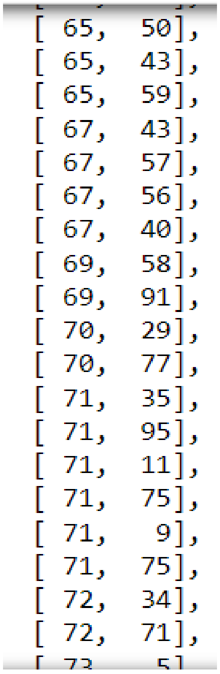
```
[ 44,   50],
[ 44,   46],
[ 46,   51],
[ 46,   46],
[ 46,   56],
[ 46,   55],
[ 47,   52],
[ 47,   59],
[ 48,   51],
[ 48,   59],
[ 48,   50],
[ 48,   48],
[ 48,   59],
[ 48,   47],
[ 49,   55],
[ 49,   42],
[ 50,   49],
[ 50,   56],
[ 54,   47],
[ 54,   54],
```



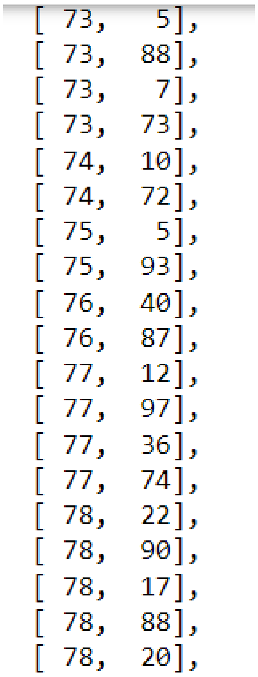
```
[ 54,  53],
[ 54,  48],
[ 54,  52],
[ 54,  42],
[ 54,  51],
[ 54,  55],
[ 54,  41],
[ 54,  44],
[ 54,  57],
[ 54,  46],
[ 57,  58],
[ 57,  55],
[ 58,  60],
[ 58,  46],
[ 59,  55],
[ 59,  41],
[ 60,  49],
[ 60,  40],
[ 60,  42],
[ 60,  52],
```



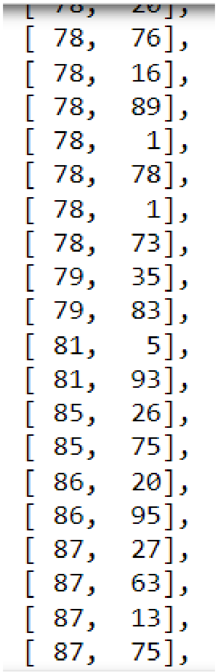
```
[ 60,  52],
[ 60,  47],
[ 60,  50],
[ 61,  42],
[ 61,  49],
[ 62,  41],
[ 62,  48],
[ 62,  59],
[ 62,  55],
[ 62,  56],
[ 62,  42],
[ 63,  50],
[ 63,  46],
[ 63,  43],
[ 63,  48],
[ 63,  52],
[ 63,  54],
[ 64,  42],
[ 64,  46],
[ 65,  48],
[ 65,  50]
```



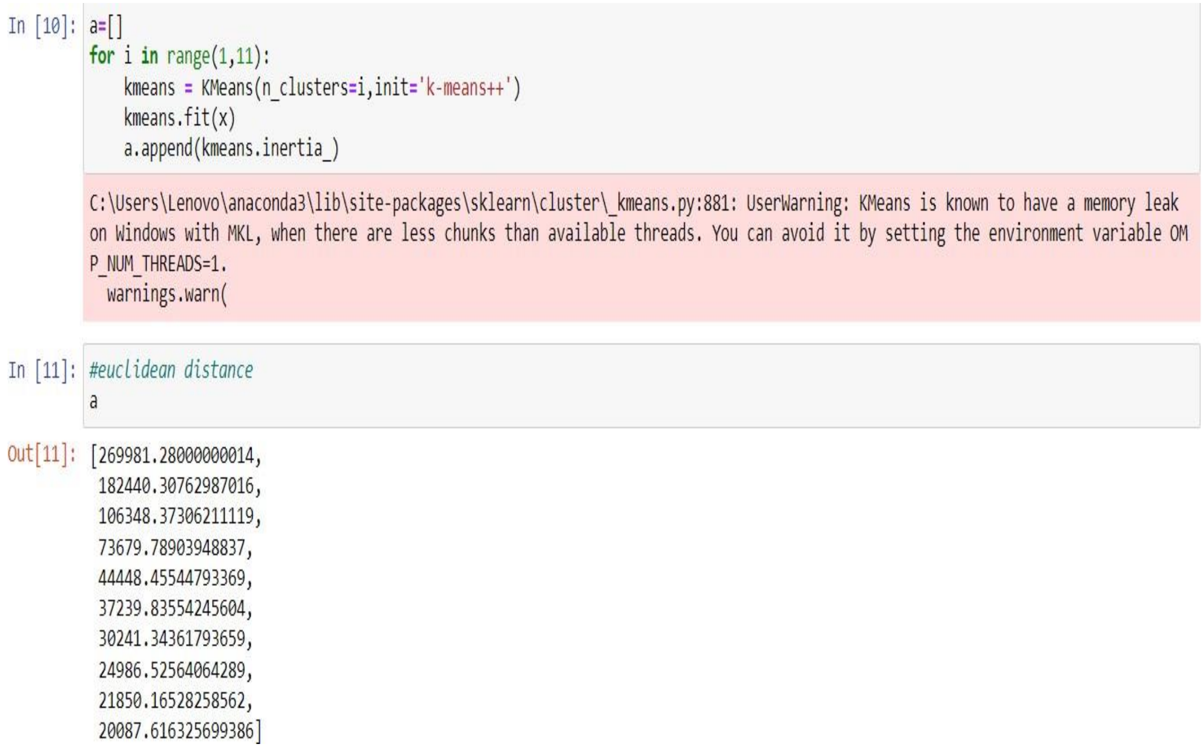
```
[ 65,  50],
[ 65,  43],
[ 65,  59],
[ 67,  43],
[ 67,  57],
[ 67,  56],
[ 67,  40],
[ 69,  58],
[ 69,  91],
[ 70,  29],
[ 70,  77],
[ 71,  35],
[ 71,  95],
[ 71,  11],
[ 71,  75],
[ 71,   9],
[ 71,  75],
[ 72,  34],
[ 72,  71],
[ 73,   5]
```



```
[ 73,   5],
[ 73,  88],
[ 73,   7],
[ 73,  73],
[ 74,  10],
[ 74,  72],
[ 75,   5],
[ 75,  93],
[ 76,  40],
[ 76,  87],
[ 77,  12],
[ 77,  97],
[ 77,  36],
[ 77,  74],
[ 78,  22],
[ 78,  90],
[ 78,  17],
[ 78,  88],
[ 78,  20],
```



```
[ 78,  20],
[ 78,  76],
[ 78,  16],
[ 78,  89],
[ 78,   1],
[ 78,  78],
[ 78,   1],
[ 78,  73],
[ 79,  35],
[ 79,  83],
[ 81,   5],
[ 81,  93],
[ 85,  26],
[ 85,  75],
[ 86,  20],
[ 86,  95],
[ 87,  27],
[ 87,  63],
[ 87,  13],
[ 87,  75],
```



In [10]:

```
a=[]
for i in range(1,11):
    kmeans = KMeans(n_clusters=i,init='k-means++')
    kmeans.fit(x)
    a.append(kmeans.inertia_)
```

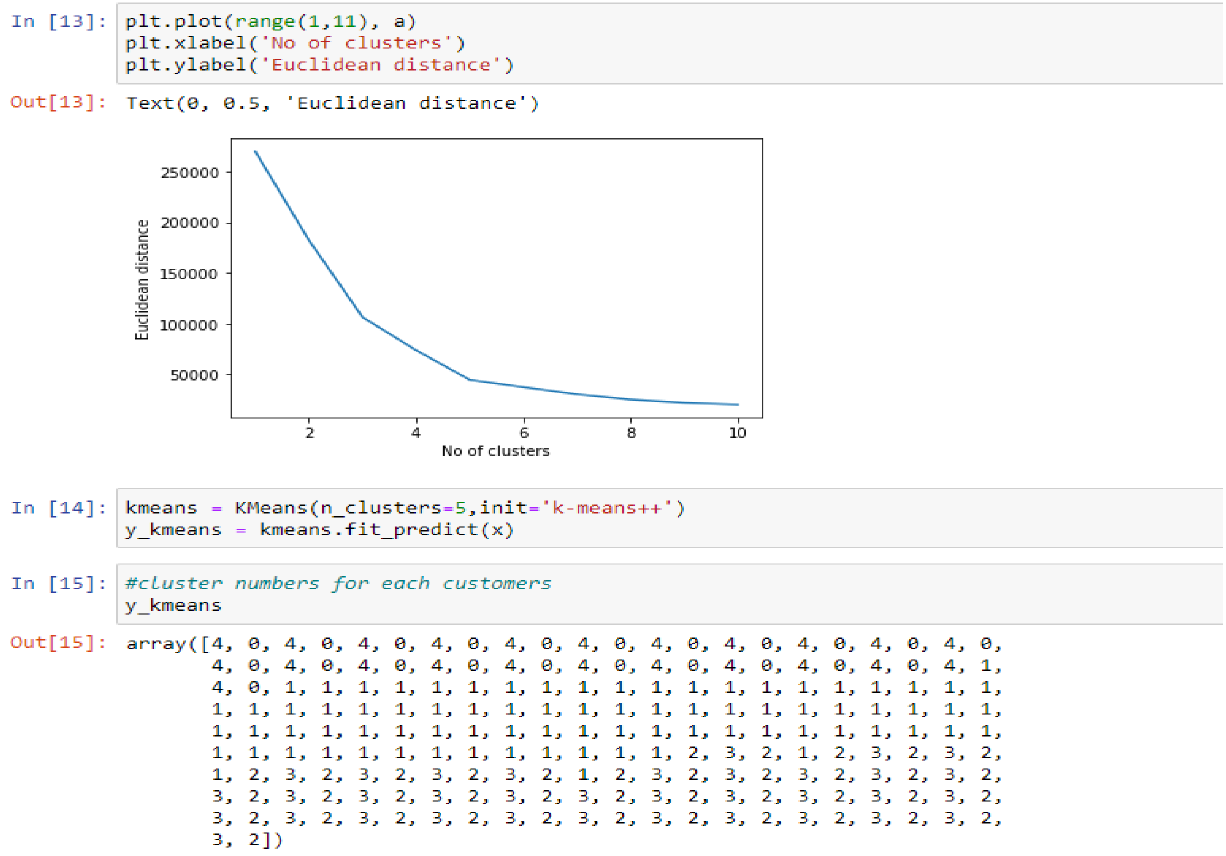
```
C:\Users\Lenovo\anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:881: UserWarning: KMeans is known to have a memory leak
on Windows with MKL, when there are less chunks than available threads. You can avoid it by setting the environment variable OM
P_NUM_THREADS=1.
  warnings.warn(
```

In [11]:

```
#euclidean distance
a
```

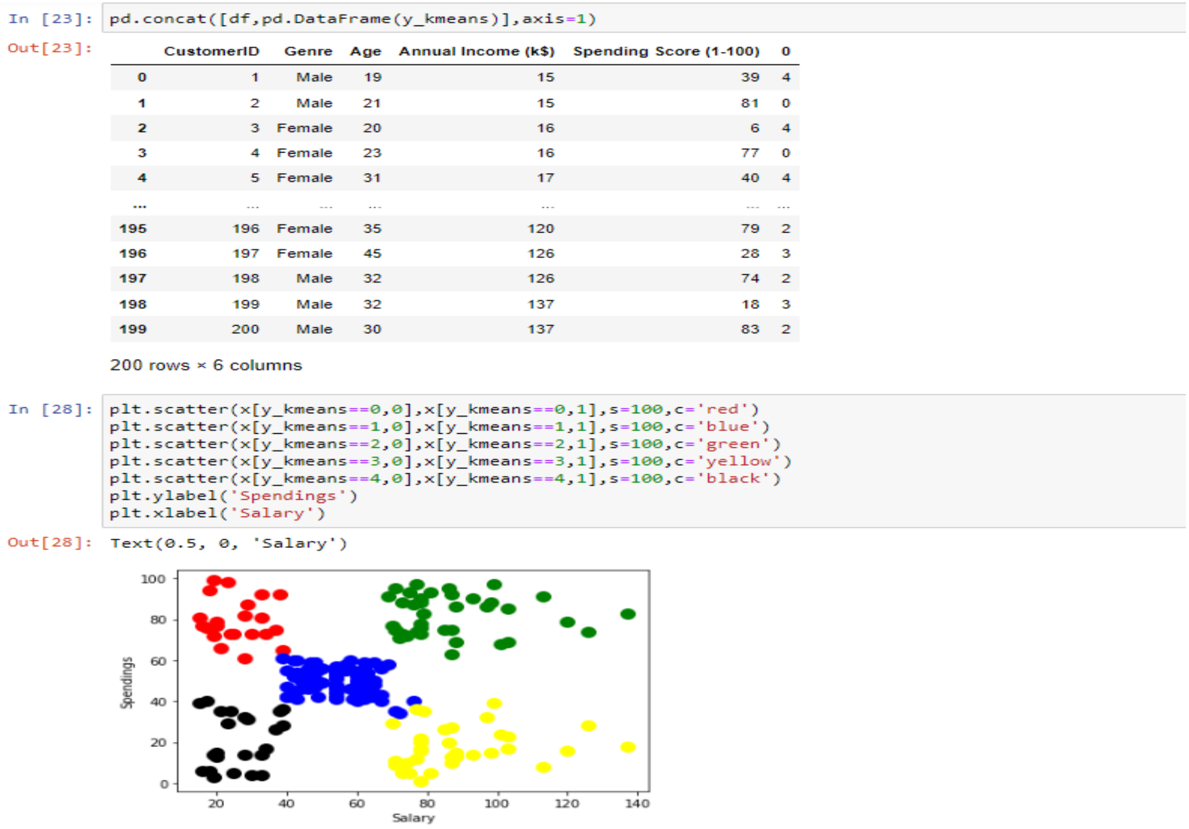
Out[11]:

```
[269981.28000000014,
 182440.30762987016,
 106348.37306211119,
 73679.78903948837,
 44448.45544793369,
 37239.83554245604,
 30241.34361793659,
 24986.52564064289,
 21850.16528258562,
 20087.616325699386]
```



```
In [13]: plt.plot(range(1,11), a)
         plt.xlabel('No of clusters')
         plt.ylabel('Euclidean distance')

Out[13]: Text(0, 0.5, 'Euclidean distance')
```

```
In [14]: kmeans = KMeans(n_clusters=5,init='k-means++')
         y_kmeans = kmeans.fit_predict(x)

In [15]: #cluster numbers for each customers
         y_kmeans

Out[15]: array([4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0,
                4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 0, 4, 1,
                4, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
                1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 2, 1, 2, 3, 2, 3, 2,
                1, 2, 3, 2, 3, 2, 3, 2, 3, 2, 1, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2,
                3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2,
                3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2, 3, 2,
                3, 2])
```



```
In [23]: pd.concat([df,pd.DataFrame(y_kmeans)],axis=1)

Out[23]:
```

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)	0
0	1	Male	19	15	39	4
1	2	Male	21	15	81	0
2	3	Female	20	16	6	4
3	4	Female	23	16	77	0
4	5	Female	31	17	40	4
...	...	...	...	...	...	...
195	196	Female	35	120	79	2
196	197	Female	45	126	28	3
197	198	Male	32	126	74	2
198	199	Male	32	137	18	3
199	200	Male	30	137	83	2

200 rows × 6 columns

```
In [28]: plt.scatter(x[y_kmeans==0,0],x[y_kmeans==0,1],s=100,c='red')
         plt.scatter(x[y_kmeans==1,0],x[y_kmeans==1,1],s=100,c='blue')
         plt.scatter(x[y_kmeans==2,0],x[y_kmeans==2,1],s=100,c='green')
         plt.scatter(x[y_kmeans==3,0],x[y_kmeans==3,1],s=100,c='yellow')
         plt.scatter(x[y_kmeans==4,0],x[y_kmeans==4,1],s=100,c='black')
         plt.ylabel('Spendings')
         plt.xlabel('Salary')

Out[28]: Text(0.5, 0, 'Salary')
```

Practical-6

Aim: Perform the Decision tree classification algorithm using a dataset.

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.preprocessing import LabelEncoder

# Load the CSV file into a DataFrame
csv_file_path = 'path_to_your_file/marksheet.csv'  # Update with your file path
df = pd.read_csv(csv_file_path)

# Preprocess the data
# Assume 'Section' is the target variable for classification
target = 'Section'
features = df.drop(columns=[target])

# Encode categorical variables
label_encoder = LabelEncoder()
df['Gender'] = label_encoder.fit_transform(df['Gender'])
df[target] = label_encoder.fit_transform(df[target])

# Define features and target
X = df.drop(columns=[target])
y = df[target]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# Train the Decision Tree model
clf = DecisionTreeClassifier(random_state=42)
clf.fit(X_train, y_train)

# Make predictions on the test set
y_pred = clf.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
conf_matrix = confusion_matrix(y_test, y_pred)
class_report = classification_report(y_test, y_pred)

# Print the results
print(f"Accuracy: {accuracy}")
print("Confusion Matrix:\n", conf_matrix)
print("Classification Report:\n", class_report)
```

Practical-7

Aim: Study and installation of the tools like PowerBI tool for data Visualization.

Introduction to Microsoft Power BI

There are several tools that make it easier for us to analyze the trajectory of business assets using data. However, Microsoft Power BI is something so unique that makes it super easy to make data-driven decisions.

It is one of the oldest, most famous, and most accessible tools for Data Visualization and business intelligence. From different sources, it collects and helps to convert into an interactive dashboard and BI reports. It is available as a Power BI Desktop, and mobile but also available for it service based on Saas. Business users use a set of services provided by it to create reports and achieve a goal. It apps help in generating reports, whereas its services help in publishing reports. It is a cloud-based service provided by Microsoft that helps viewers to visualize data, and this process is called Data visualization with it.

It helps in sorting, analyzing, and comparing data more frequently and quickly. It is compatible with various resources like Microsoft Excel, Access Database, SQL Server, which helps make the best choice for data analysts.

Installation process of Power BI

Two Ways to install Power BI and Get it up and Running:

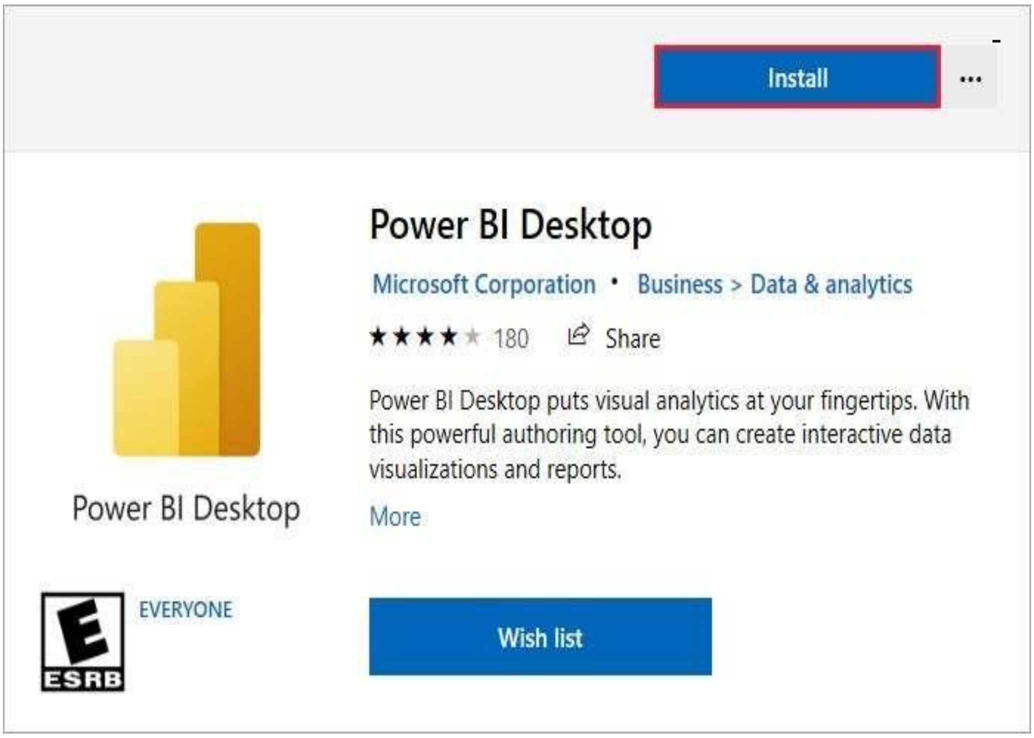
Either of the two approaches gets the latest version of Power BI Desktop onto your computer.

1. **Install as an app from the Microsoft Store**.

There are a few ways to access the most recent version of Power BI Desktop from the Microsoft Store. Use one of the following options to open the Power BI Desktop page of the Microsoft Store:

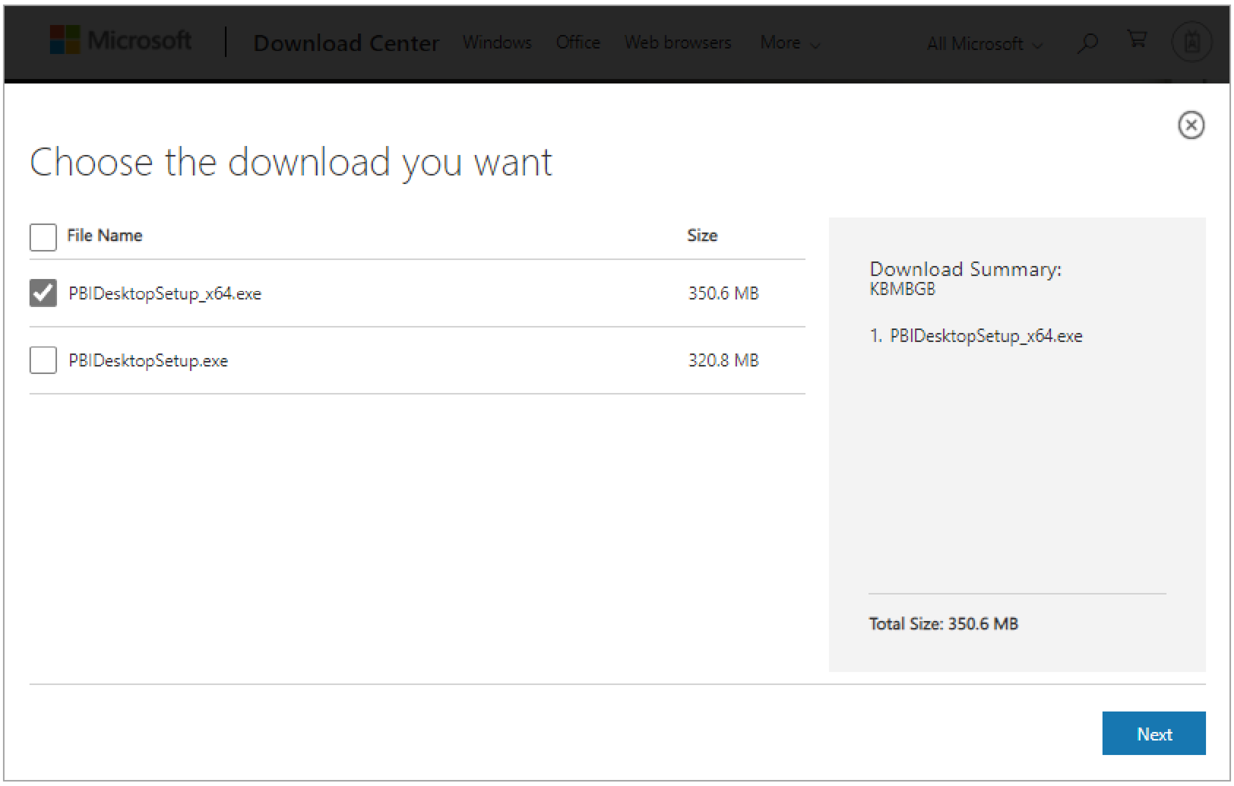
1. Open a browser and go directly to the Power BI Desktop page of the Microsoft Store.
2. From the Power BI service, select the Download icon in the upper right corner and then choose Power BI Desktop.
3. Go to the Power BI Desktop product page and select Download Free.

4. After you've landed on the Power BI Desktop page of the Microsoft Store, select Install.


Install
...
Power BI Desktop
Microsoft Corporation • Business > Data & analytics
★★★★★ 180
Share
Power BI Desktop puts visual analytics at your fingertips. With this powerful authoring tool, you can create interactive data visualizations and reports.
More
Power BI Desktop
E
ESRB
EVERYONE
Wish list

2. **Download directly as an executable; you download and install it on your computer.**

Select Download from the Download Center to download the Power BI Desktop executable from the Download Center page. Then, specify a 32-bit or 64-bit installation file to download.


Microsoft | Download Center Windows Office Web browsers More All Microsoft

Choose the download you want

□ File Name	Size
☑ PBIDesktopSetup_x64.exe	350.6 MB
□ PBIDesktopSetup.exe	320.8 MB

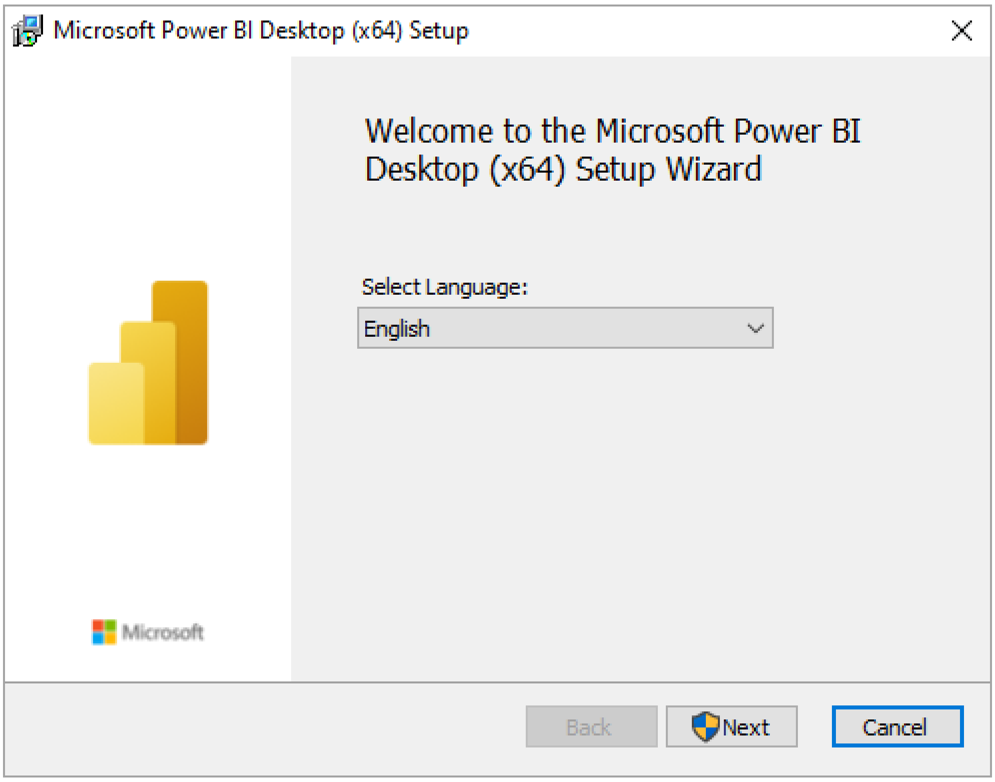
Download Summary:
KBMBGB

1. PBIDesktopSetup_x64.exe

Total Size: 350.6 MB

Install Power BI Desktop after downloading it

You're prompted to run the installation file after downloading it.



Welcome to the Microsoft Power BI Desktop (x64) Setup Wizard

Select Language: English

And that’s it; you are good to go.

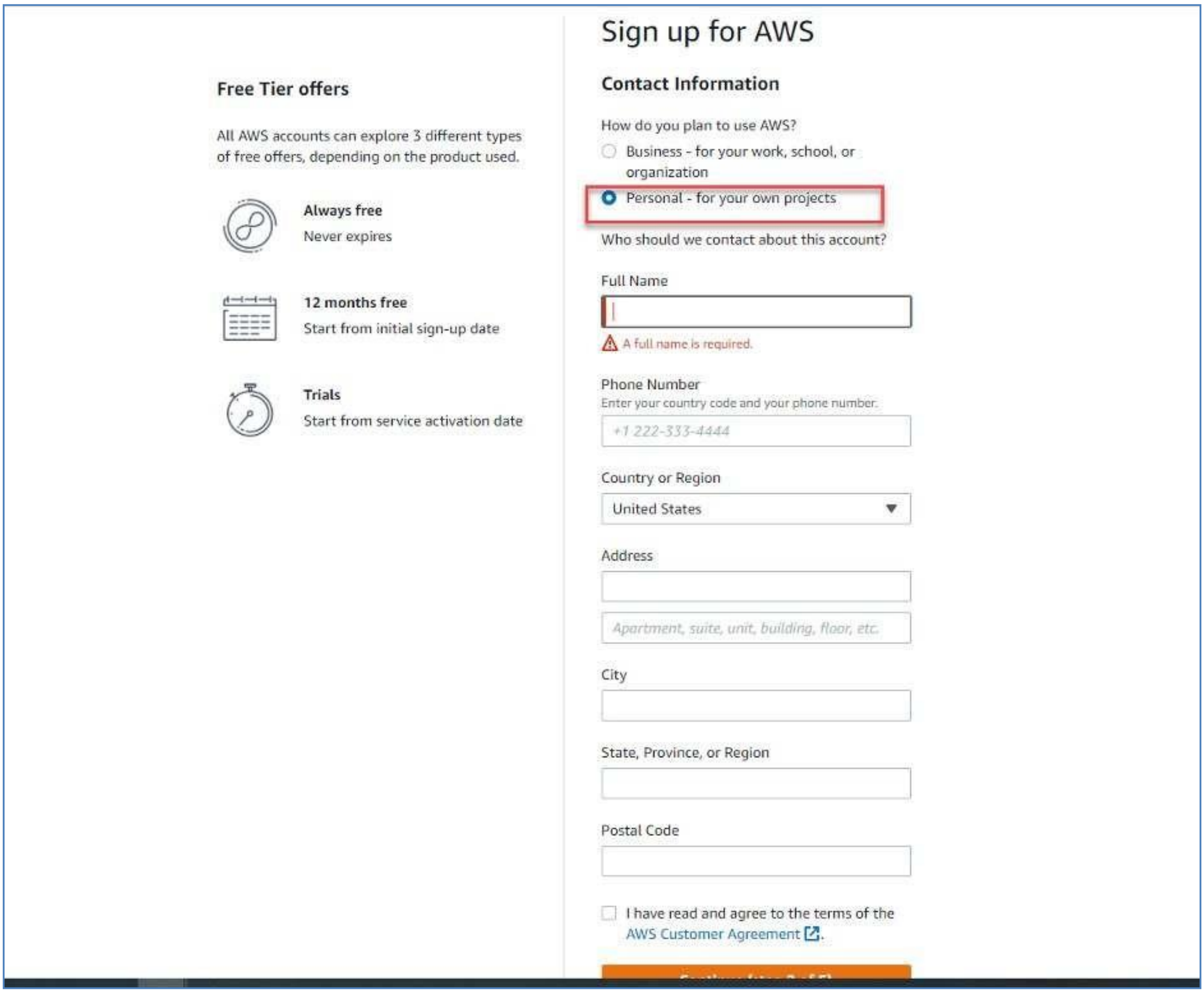
Note: Windows Users can ignore the instructions beyond this point. The instructions ahead are specifically for MAC users.

Installing Instructions for MAC

Since Power BI is not directly available for mac users, we will use an AWS container to runan instance of windows ten where you can download power BI and install it for use.

1. Create an AWS account using this URL https://portal.aws.amazon.com/billing/signup#/start

2. Select Personal when asked how you plan to use this account



Free Tier offers

All AWS accounts can explore 3 different types of free offers, depending on the product used.

Always free
Never expires

12 months free
Start from initial sign-up date

Trials
Start from service activation date

Sign up for AWS

Contact Information

How do you plan to use AWS?

○ Business - for your work, school, or organization

● Personal - for your own projects

Who should we contact about this account?

Full Name

A full name is required.

Phone Number
Enter your country code and your phone number.
+1 222-333-4444

Country or Region
United States

Address

Apartment, suite, unit, building, floor, etc.

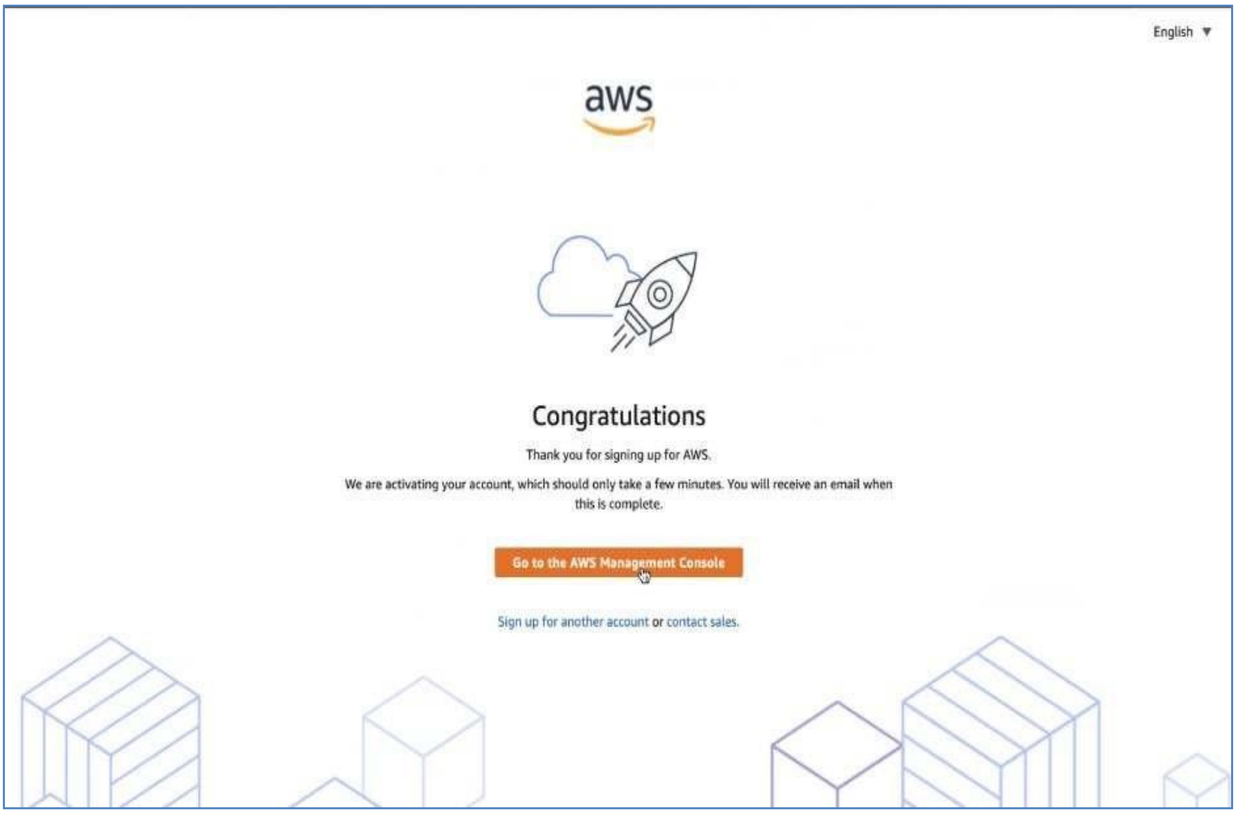
City

State, Province, or Region

Postal Code

□ I have read and agree to the terms of the AWS Customer Agreement.

3. After completing all the steps, you will be asked to enter billing information, which is a mandatory step.



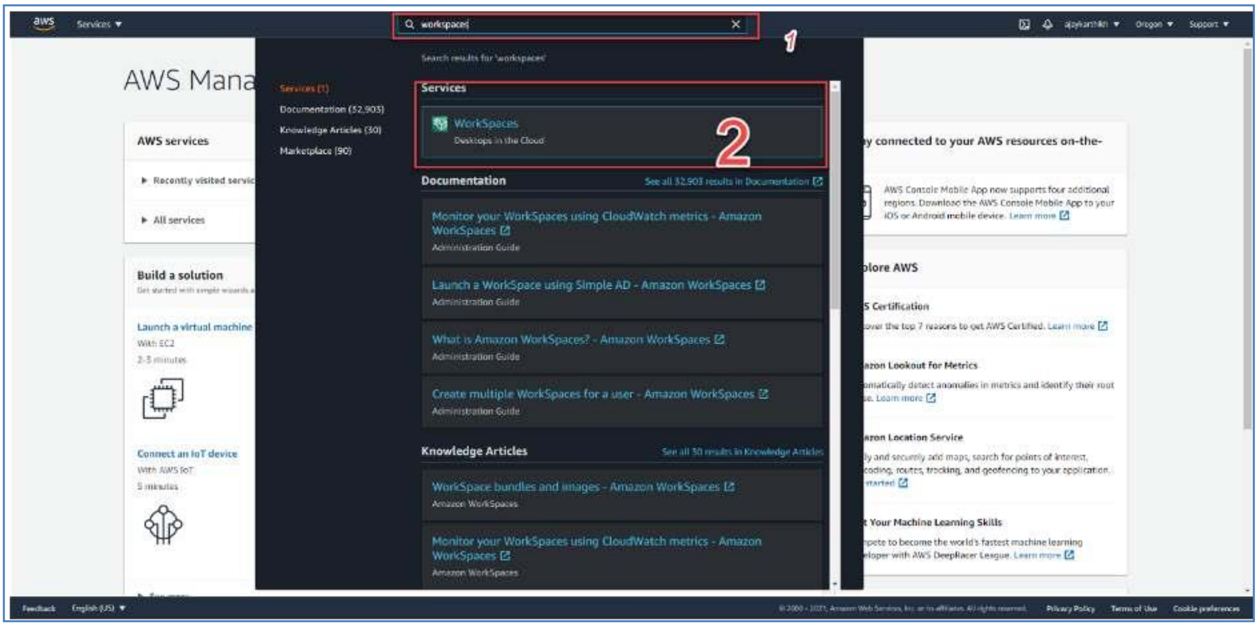
Congratulations

Thank you for signing up for AWS.

We are activating your account, which should only take a few minutes. You will receive an email when this is complete.

Go to the AWS Management Console
Sign up for another account or contact sales.

4. Login into your AWS console, search for "workspaces," and click on the Workspacesicon as shown below.



Search results for 'workspaces'

Services

WorkSpaces
Desktops in the Cloud

Documentation — See all 32,903 results in Documentation

Monitor your WorkSpaces using CloudWatch metrics - Amazon WorkSpaces
Administration Guide

Launch a WorkSpace using Simple AD - Amazon WorkSpaces
Administration Guide

What is Amazon WorkSpaces? - Amazon WorkSpaces
Administration Guide

Create multiple WorkSpaces for a user - Amazon WorkSpaces
Administration Guide

Knowledge Articles — See all 30 results in Knowledge Articles

WorkSpace bundles and images - Amazon WorkSpaces
Amazon WorkSpaces

Monitor your WorkSpaces using CloudWatch metrics - Amazon WorkSpaces
Amazon WorkSpaces

y connected to your AWS resources on-the-

AWS Console Mobile App now supports four additional regions. Download the AWS Console Mobile App to your iOS or Android mobile device. Learn more

plore AWS

S Certification
over the top 7 reasons to get AWS Certified. Learn more

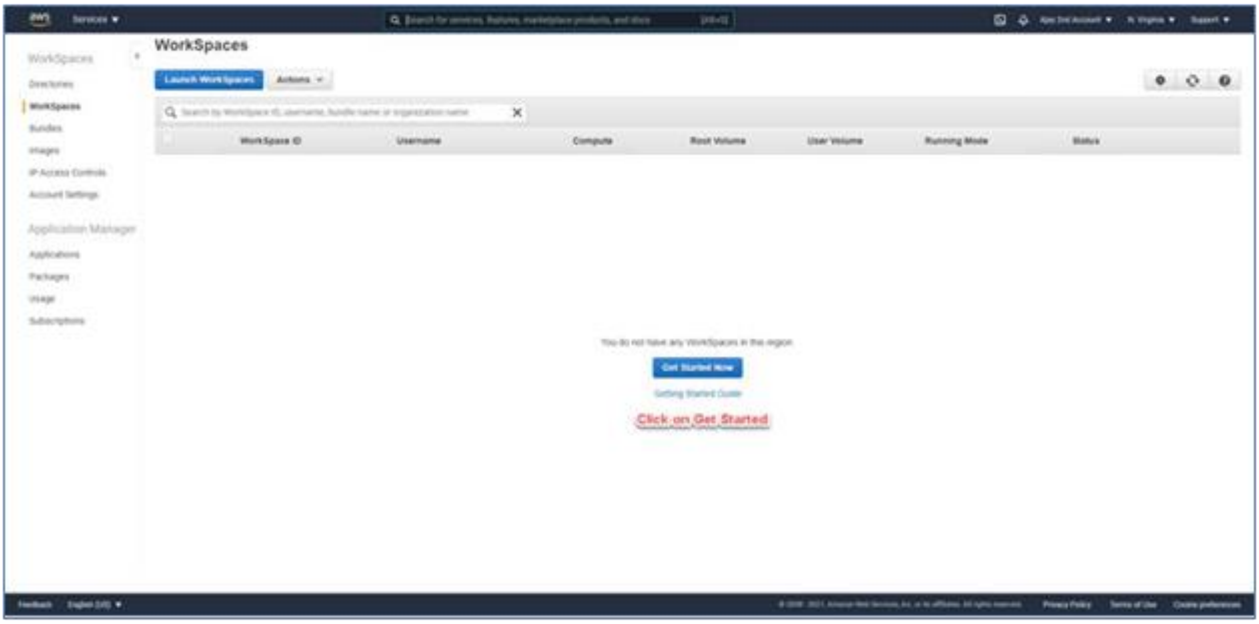
azon Lookout for Metrics
omatically detect anomalies in metrics and identify their root se. Learn more

azon Location Service
ly and securely add maps, search for points of interest, coding, routes, tracking, and geofencing to your application. tarted

t Your Machine Learning Skills
pete to become the world's fastest machine learning eloper with AWS DeepRacer League. Learn more

© 2000 - 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

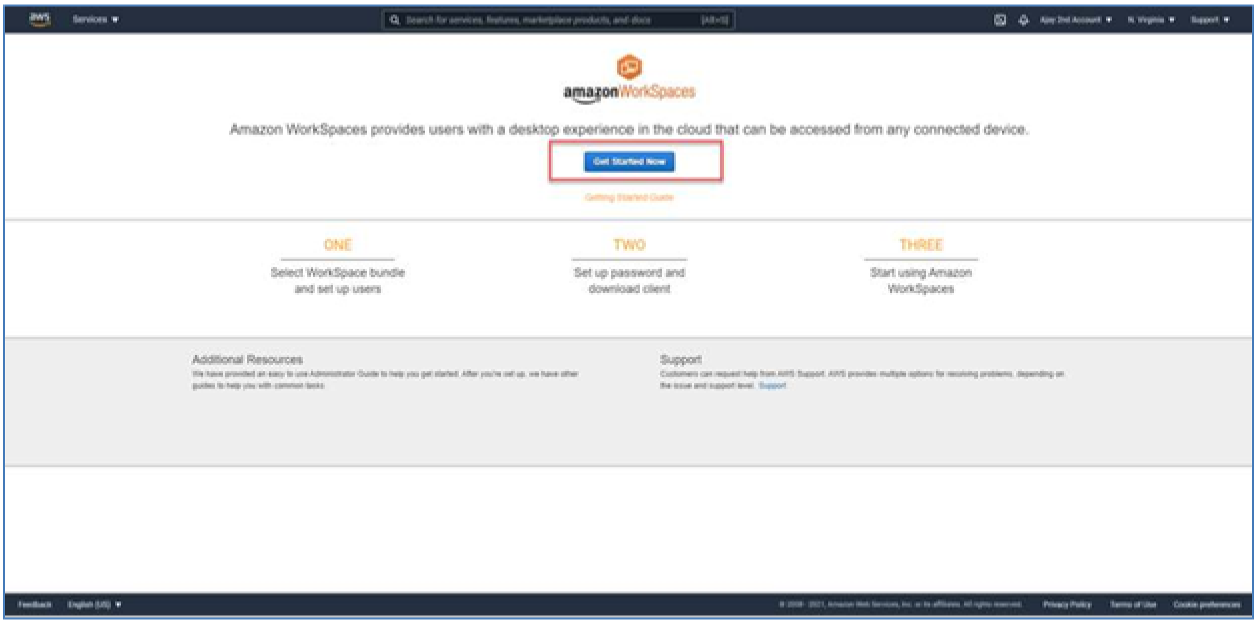
5. Here you should be able to see a Launch Workspace.



WorkSpaces

You do not have any WorkSpaces in this region.

Click on Get Started



Amazon WorkSpaces provides users with a desktop experience in the cloud that can be accessed from any connected device.

ONE: Select WorkSpace bundle and set up users

TWO: Set up password and download client

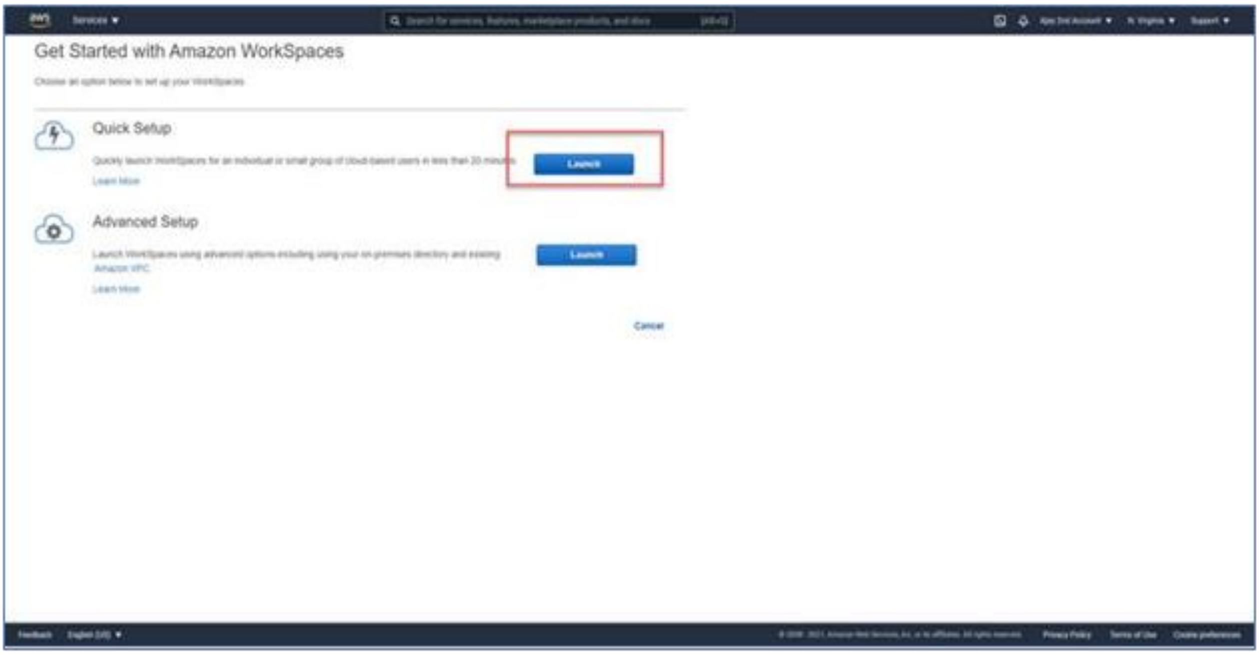
THREE: Start using Amazon WorkSpaces

Additional Resources

We have provided an easy to use Administrator Guide to help you get started. After you're set up, we have other guides to help you with common tasks.

Support

Customers can request help from AWS Support. AWS provides multiple options for resolving problems, depending on the issue and support level. Support



Get Started with Amazon WorkSpaces

Choose an option below to set up your WorkSpaces

Quick Setup

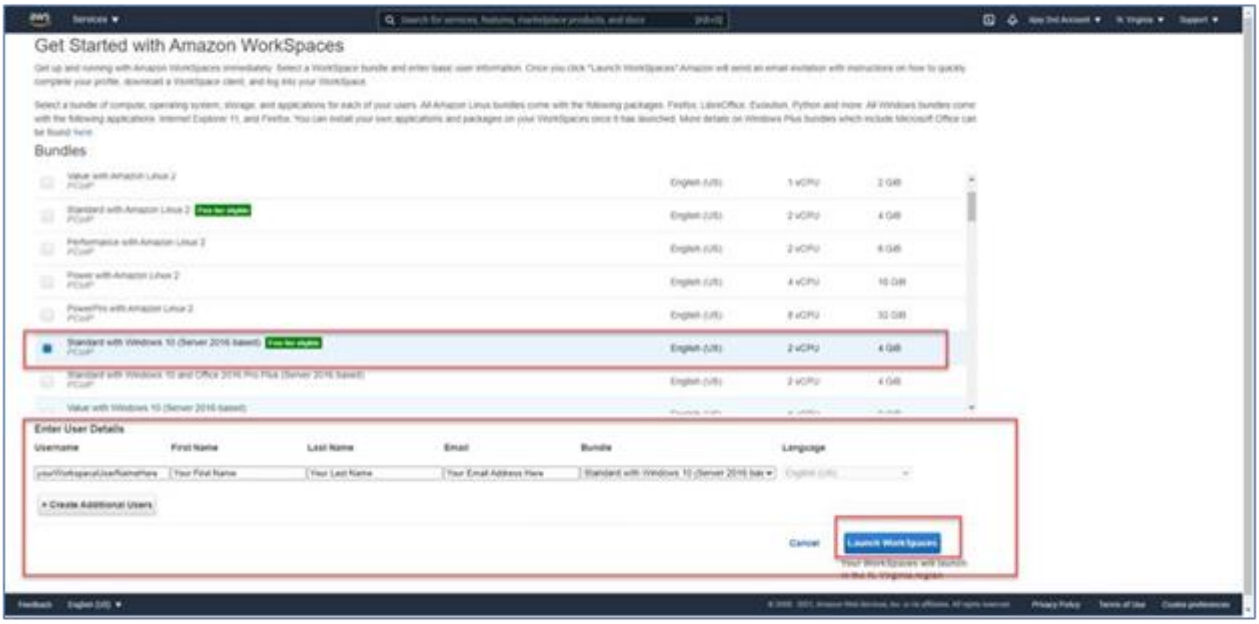
Quickly launch WorkSpaces for an individual or small group of cloud-based users in less than 20 minutes

Learn More

Advanced Setup

Launch WorkSpaces using advanced options including using your on-premises directory and existing Amazon VPC

Learn More



Get Started with Amazon WorkSpaces

Get up and running with Amazon WorkSpaces immediately. Select a WorkSpace bundle and enter basic user information. Once you click "Launch WorkSpaces" Amazon will send an email invitation with instructions on how to quickly complete your profile, download a WorkSpace client, and log into your WorkSpace.

Select a bundle of compute, operating system, storage, and applications for each of your users. All Amazon Linux bundles come with the following packages: Firefox, LibreOffice, Evolution, Python and more. All Windows bundles come with the following applications: Internet Explorer 11, and Firefox. You can install your own applications and packages on your WorkSpaces once it has launched. More details on Windows Plus bundles which include Microsoft Office can be found here.

Bundles

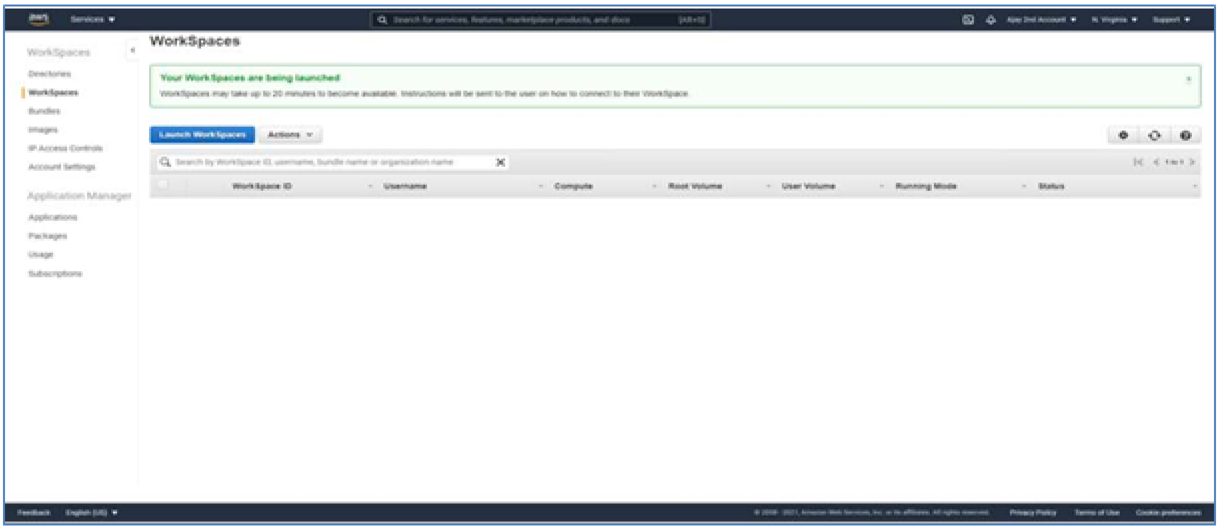
○	Value with Amazon Linux 2 PCoIP	English (US)	1 vCPU	2 GiB
○	Standard with Amazon Linux 2 Free tier eligible PCoIP	English (US)	2 vCPU	4 GiB
○	Performance with Amazon Linux 2 PCoIP	English (US)	2 vCPU	8 GiB
○	Power with Amazon Linux 2 PCoIP	English (US)	4 vCPU	16 GiB
○	PowerPro with Amazon Linux 2 PCoIP	English (US)	8 vCPU	32 GiB
●	Standard with Windows 10 (Server 2016 based) Free tier eligible PCoIP	English (US)	2 vCPU	4 GiB
○	Standard with Windows 10 and Office 2016 Pro Plus (Server 2016 based) PCoIP	English (US)	2 vCPU	4 GiB
○	Value with Windows 10 (Server 2016 based)	[illegible]	[illegible]	[illegible]

Enter User Details

Username	First Name	Last Name	Email	Bundle	Language
yourWorkspaceUserNameHere	Your First Name	Your Last Name	Your Email Address Here	Standard with Windows 10 (Server 2016 bas ▾	English (US) ▾

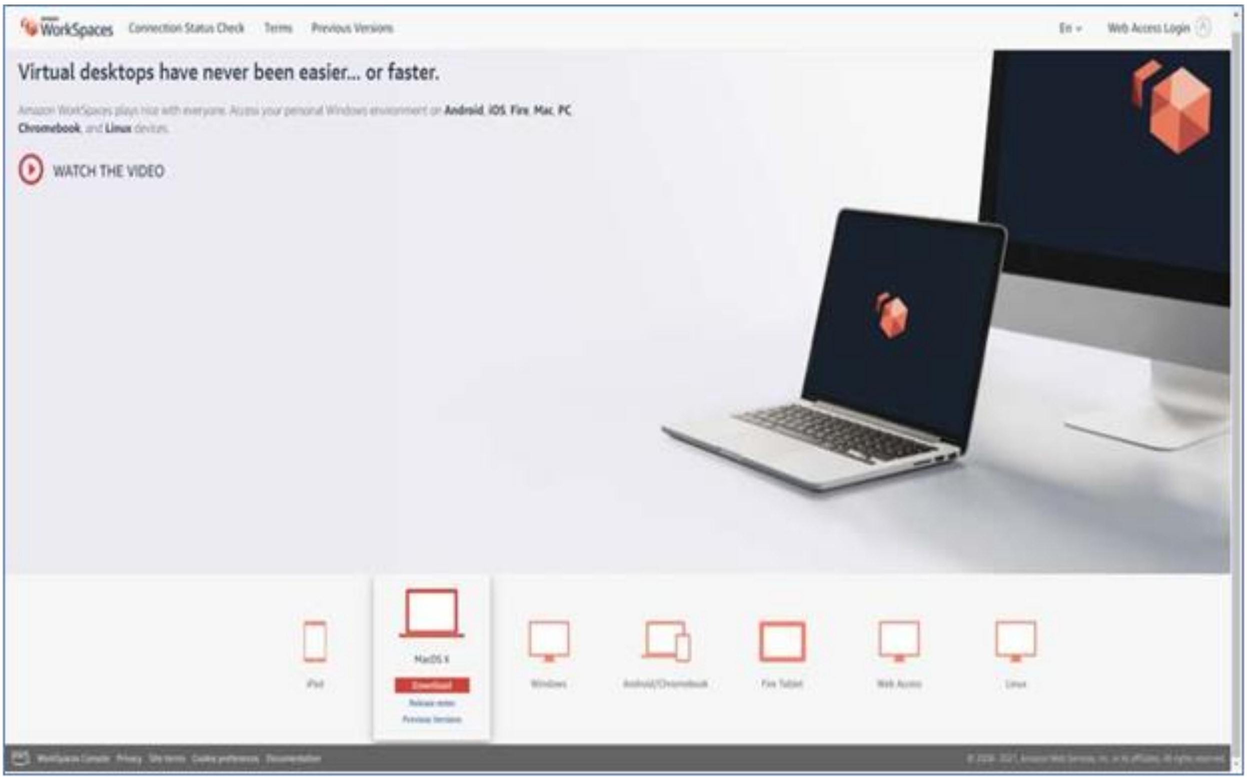
Your WorkSpaces will launch in the N. Virginia region

When you click launch, you will be prompted to view the workspace console; clicking on this button will take you to the console where it says that it will take 20 min to set the workspace; typically, it takes 15-30 mining, my experience.



Your WorkSpaces are being launched
WorkSpaces may take up to 20 minutes to become available. Instructions will be sent to the user on how to connect to their WorkSpace.

14. Now in this waiting time, download the appropriate Workspace RD client from this link - https://clients.amazonworkspaces.com/?refid=ps_a134p000006vmxaaai&trkcampaign=acq_paid_search_brand

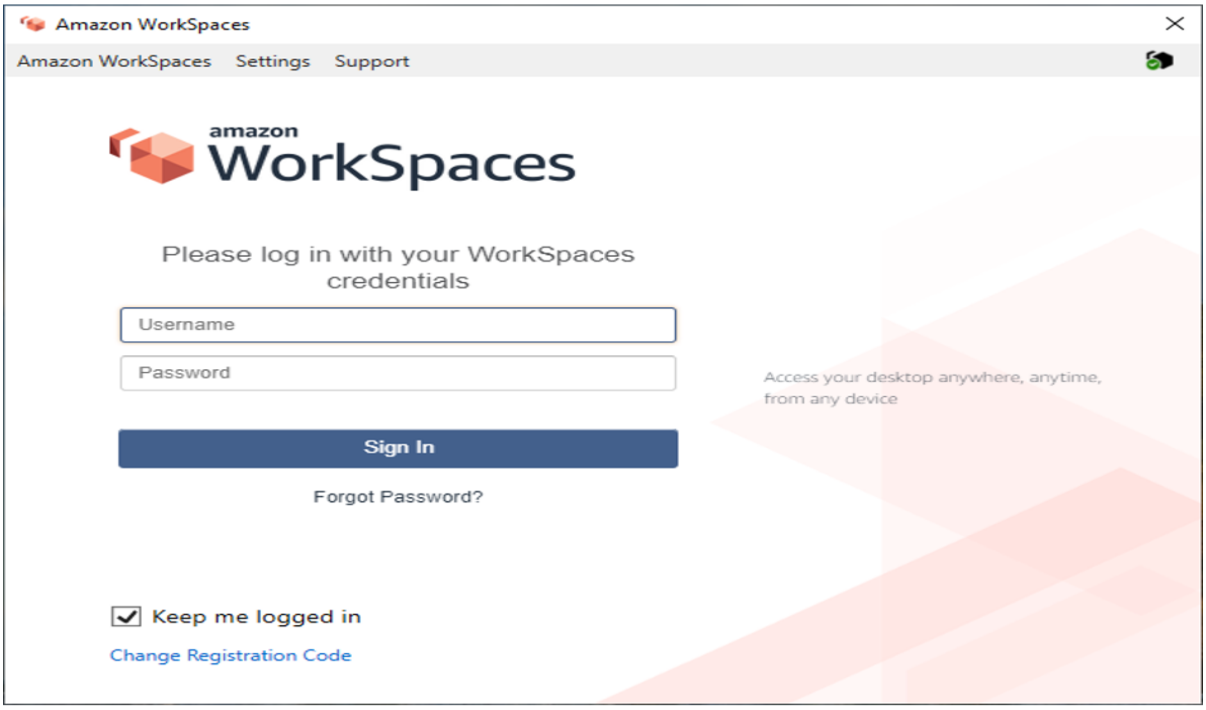


Virtual desktops have never been easier... or faster.

Amazon WorkSpaces plays nice with everyone. Access your personal Windows environment on Android, iOS, Fire, Mac, PC Chromebook, and Linux devices.

15. Now Install this and wait for your Workspace to finish setting up

16. Start the Amazon Workspace client on your Mac, and you should see something like the below screenshot



Please log in with your WorkSpaces credentials

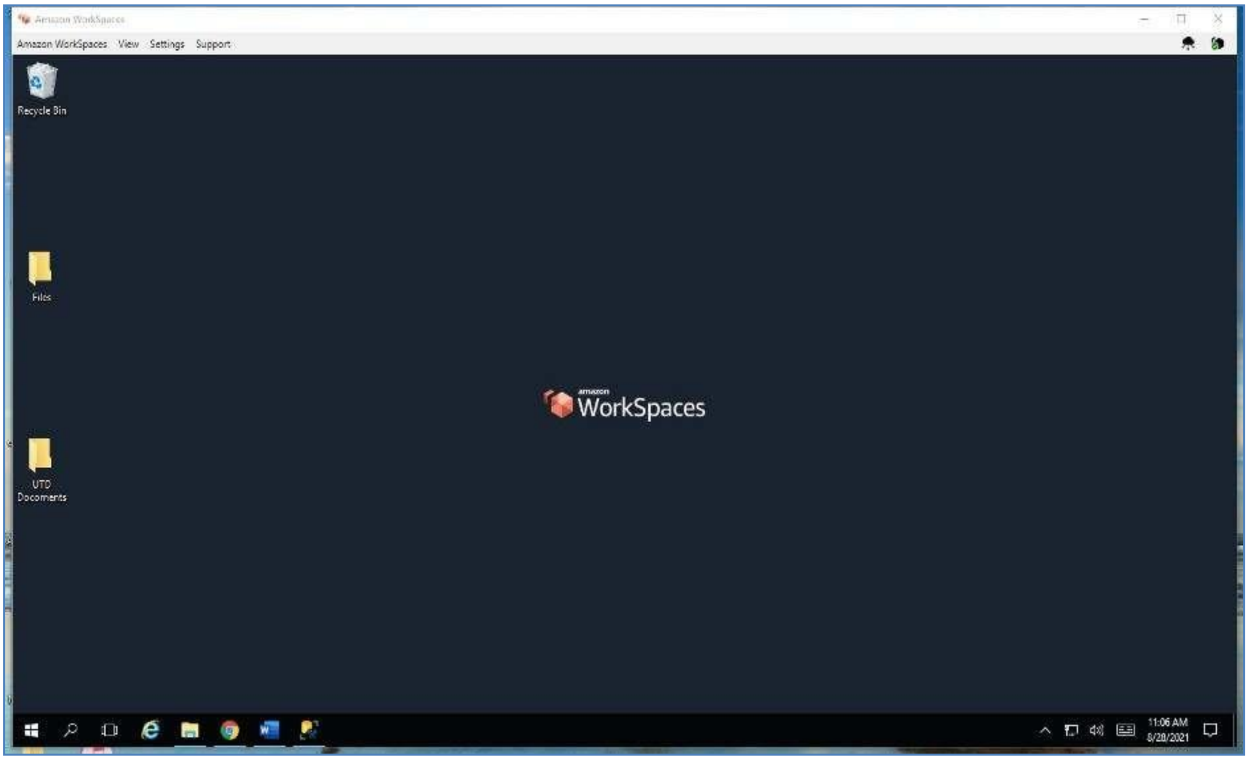
Username

Password

☑ Keep me logged in

Access your desktop anywhere, anytime, from any device

14. Remember that you did not create a password in step 11. It would help if you were gettingan email to create a password. Use that password and the username used in step 11 here.
15. After successful login, you should see something like this.



20. Select Download from the Download Center to download the Power BI Desktop executable from the Download Center page. Then, specify a 32-bit or 64-bit installation file to download.
21. Follow all the steps followed for Power BI installation on Windows.

Practical-8

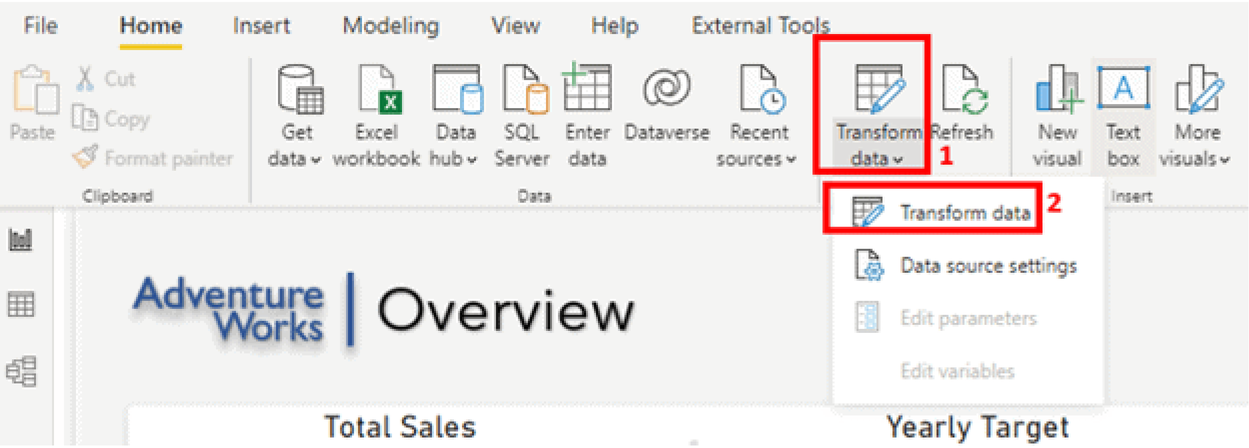
Aim: Load a dataset from different sources in PowerBI and apply transformations to it.

Step 1: Open the Power BI report in Power BI Desktop.

First, you need to open the Power BI report that you want to update in Power BI Desktop.

Step 2: Click **Transform data**.

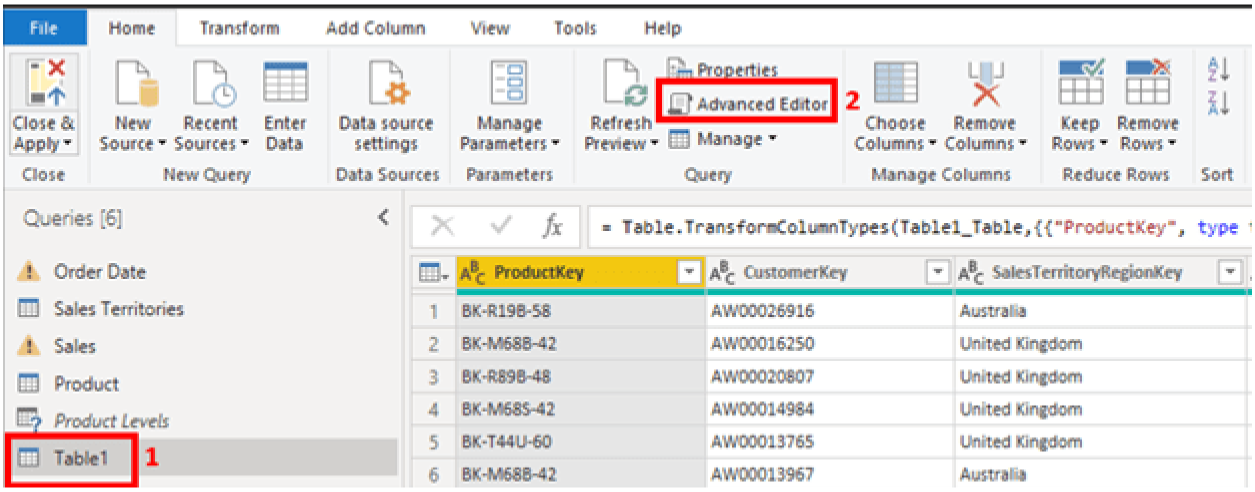
Navigate to the “Data” tab, click the **Transform data** drop-down, and then **Transform data** to get to Power Query.



Step 3: Click on **New Source** and connect to the new data source. Confirm the column names in Table 1 (new data source) are the same as that of the Sales table (current data source).

Step 4: Click on **Advanced Editor**.

Select **Table 1**, navigate to the “View” tab, and click on **Advanced Editor**.

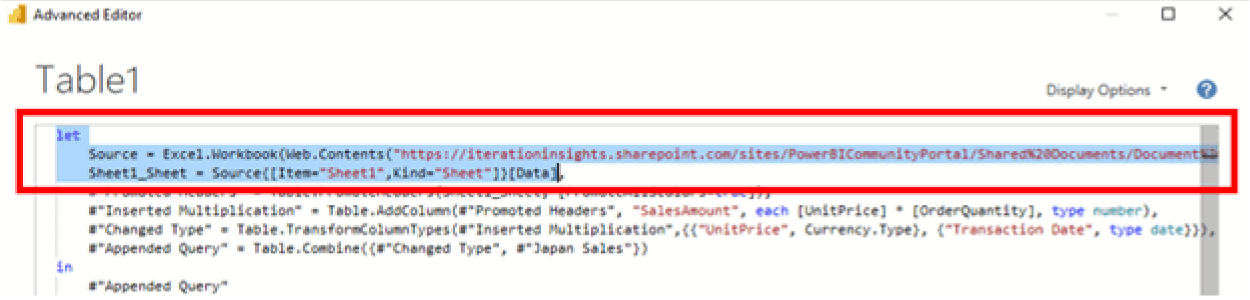


```
= Table.TransformColumnTypes(Table1_Table,{{"ProductKey", type
```

	ProductKey	CustomerKey	SalesTerritoryRegionKey
1	BK-R19B-58	AW00026916	Australia
2	BK-M68B-42	AW00016250	United Kingdom
3	BK-R89B-48	AW00020807	United Kingdom
4	BK-M68S-42	AW00014984	United Kingdom
5	BK-T44U-60	AW00013765	United Kingdom
6	BK-M68B-42	AW00013967	Australia

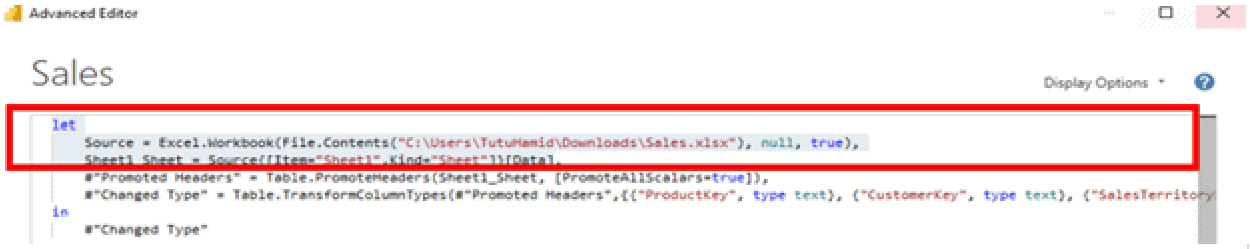
Step 5: Copy the M code up to [data].

Locate the M code that specifies the new data source and copy from beginning up to [data].

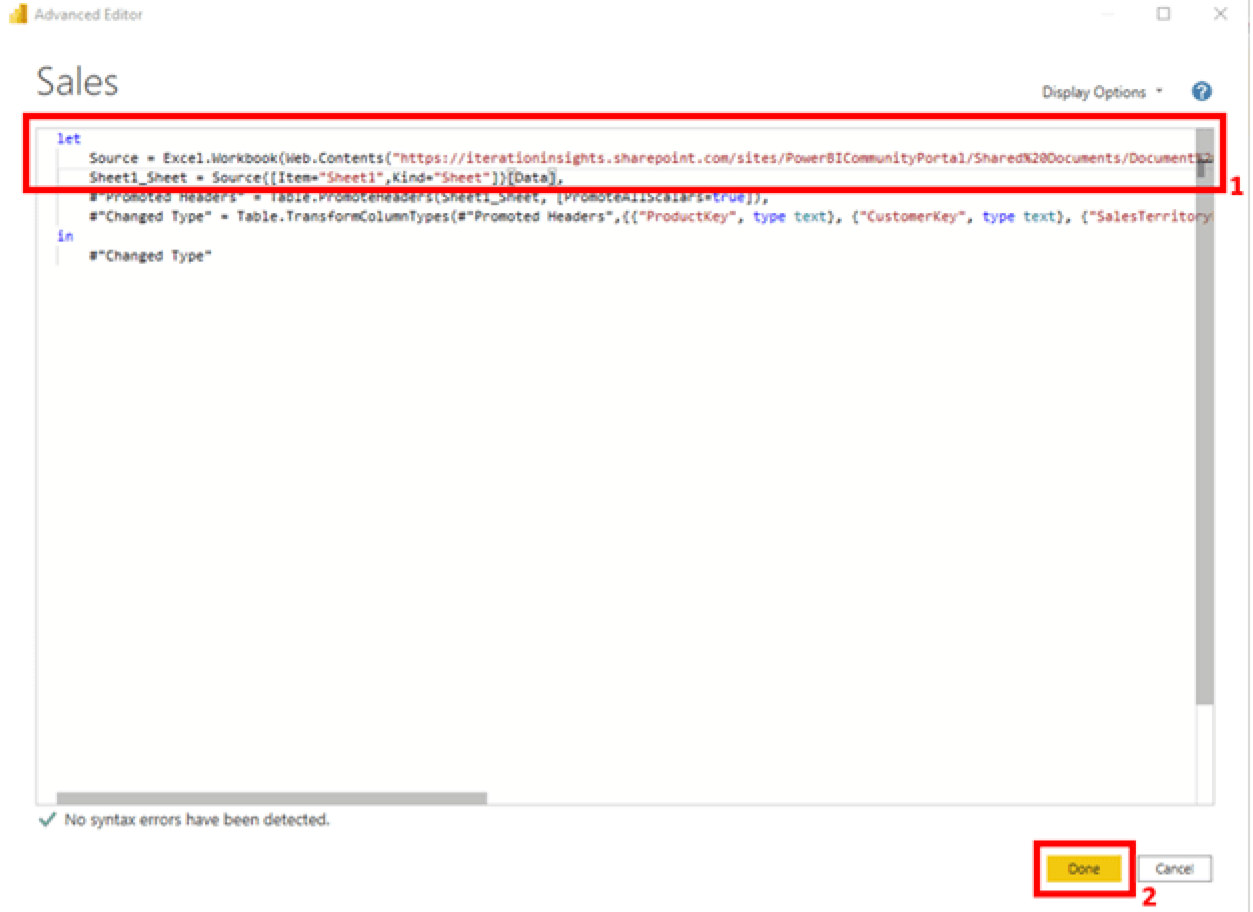


```
let
    Source = Excel.Workbook(Web.Contents("https://iterationinsights.sharepoint.com/sites/PowerBICommunityPortal/Shared%20Documents/Document%2
    Sheet1_Sheet = Source{[Item="Sheet1",Kind="Sheet"]}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(Sheet1_Sheet, [PromoteAllScalars=true]),
    #"Inserted Multiplication" = Table.AddColumn(#"Promoted Headers", "SalesAmount", each [UnitPrice] * [OrderQuantity], type number),
    #"Changed Type" = Table.TransformColumnTypes(#"Inserted Multiplication",{{"UnitPrice", Currency.Type}, {"Transaction Date", type date}}),
    #"Appended Query" = Table.Combine({#"Changed Type", #"Japan Sales"})
in
    #"Appended Query"
```

Step 6: Select Sales table and click Advance Editor. Replace the M code up to [data] with the code you copied from Table1 and click done.



```
let
    Source = Excel.Workbook(File.Contents("C:\Users\TutuHamid\Downloads\Sales.xlsx"), null, true),
    Sheet1_Sheet = Source{[Item="Sheet1",Kind="Sheet"]}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(Sheet1_Sheet, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"ProductKey", type text}, {"CustomerKey", type text}, {"SalesTerritory
in
    #"Changed Type"
```



```
let
    Source = Excel.Workbook(Web.Contents("https://iterationinsights.sharepoint.com/sites/PowerBICommunityPortal/Shared%20Documents/Document%2
    Sheet1_Sheet = Source{[Item="Sheet1",Kind="Sheet"]}[Data],
    #"Promoted Headers" = Table.PromoteHeaders(Sheet1_Sheet, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"ProductKey", type text}, {"CustomerKey", type text}, {"SalesTerritory
in
    #"Changed Type"
```

After making the necessary changes, the user should review the M code and ensure it accurately reflects the new data source.

Step 7: Delete Table1 and click **Close & Apply**.

Practical-9

Aim: Study and Plot various graphs for Data Visualization on PowerBI.

Scatter charts

Scatter charts work well in many scenarios:

- Show relationships between two numerical values.
- Plot two groups of numbers as one series of x and y coordinates.
- Display worksheet data with pairs or grouped sets of values.
- Show patterns in large sets of data.
- Compare large amounts of data points irrespective of time measurements.
- Convert horizontal axis into logarithmic scale.
- Substitute for line charts to enable changing horizontal axis scale.

Bubble charts

You can use a bubble chart in many of the same scenarios as a scatter chart. Here are some of the other ways you can use bubble charts:

- Visually emphasize value differences with variable bubble size.
- Support scenarios with three data series that each has sets of values.
- Present financial data in a visual rather than numerical form.
- Display data with quadrants.

Dot plot charts

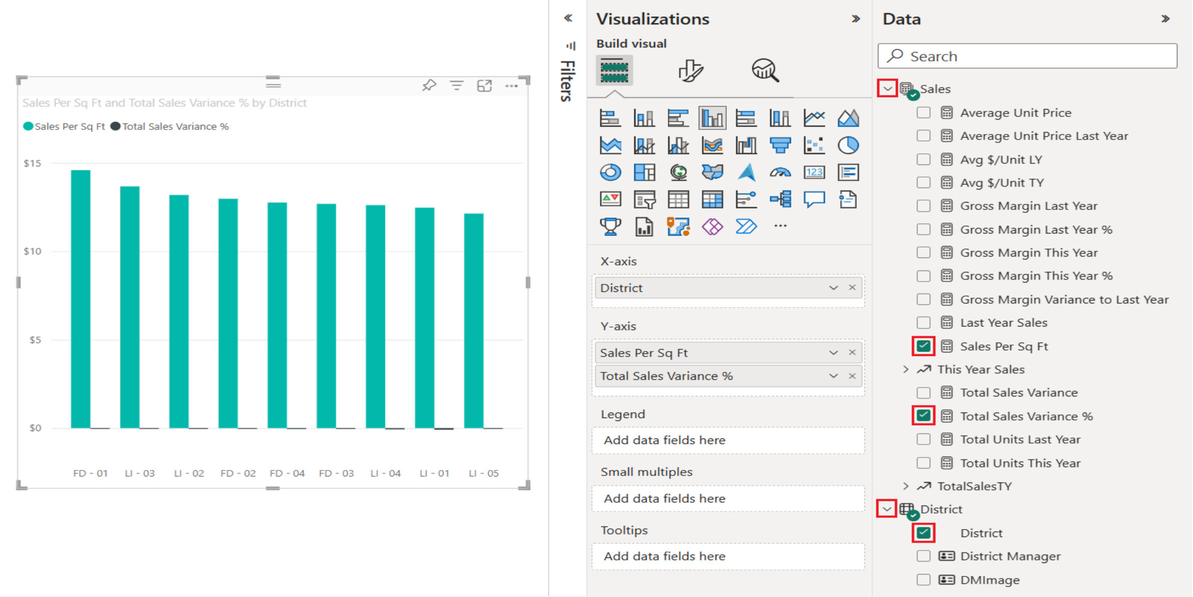
Use cases for the dot plot chart are similar to the scenarios described for scatter and bubble charts. The primary advantage of dot plot charts is the ability to include categorical data along the horizontal axis.

Create a scatter chart

1. On the **Data** pane, select three fields:

- Expand **Sales** and select the **Sales Per Sq Ft** and **Total Sales Variance %** checkboxes.
- Expand **District** and select the **District** checkbox.

By default, Power BI creates a clustered column chart to display the data. On the **Visualizations** pane, the **District** field is assigned to the **X-axis** and the other two fields are assigned to the **Y-axis**.



X-axis
District

Y-axis
Sales Per Sq Ft
Total Sales Variance %

Legend
Add data fields here

Small multiples
Add data fields here

Tooltips
Add data fields here

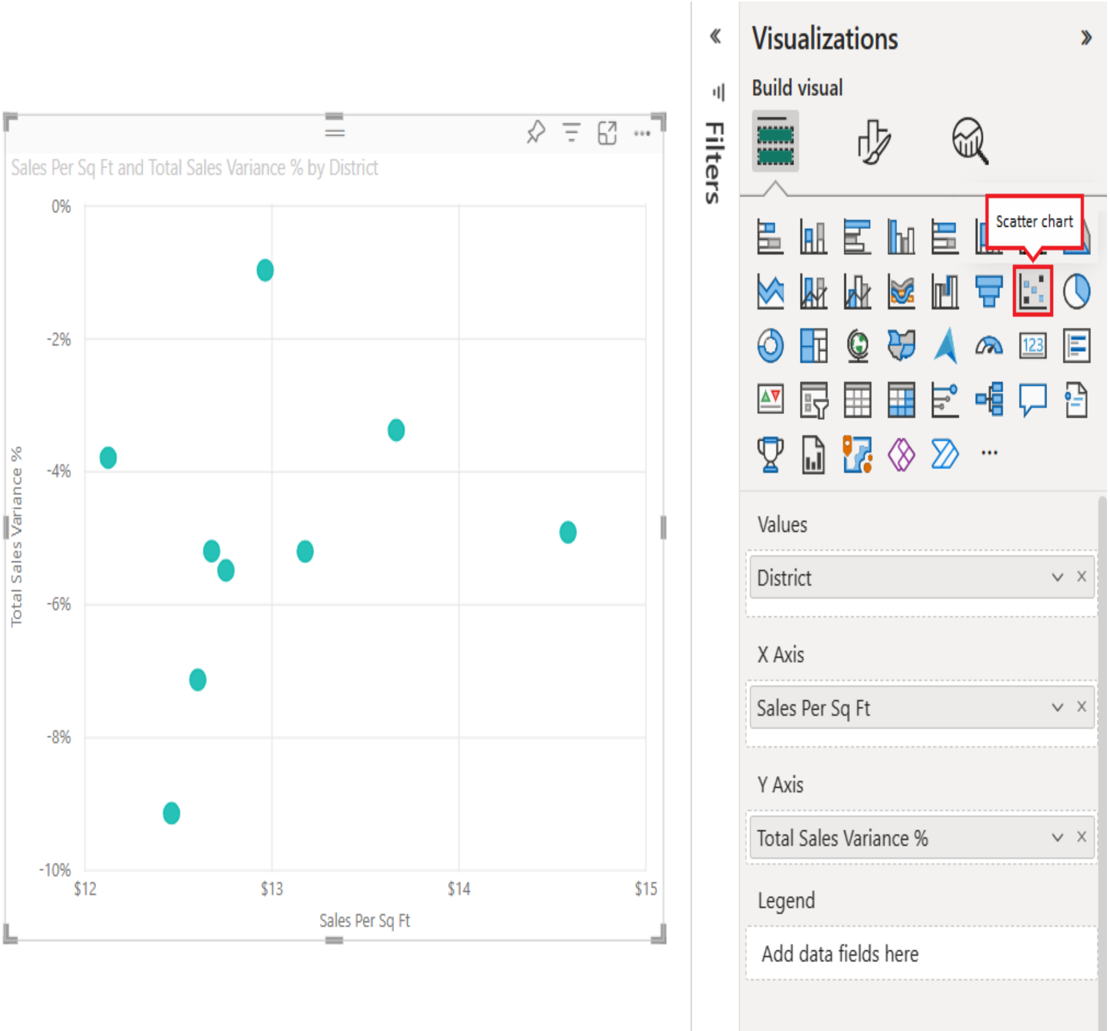
Sales
- □ Average Unit Price
- □ Average Unit Price Last Year
- □ Avg $/Unit LY
- □ Avg $/Unit TY
- □ Gross Margin Last Year
- □ Gross Margin Last Year %
- □ Gross Margin This Year
- □ Gross Margin This Year %
- □ Gross Margin Variance to Last Year
- □ Last Year Sales
- ☑ Sales Per Sq Ft
- This Year Sales
- □ Total Sales Variance
- ☑ Total Sales Variance %
- □ Total Units Last Year
- □ Total Units This Year
- TotalSalesTY

District
- ☑ District
- □ District Manager
- □ DMImage

You can now convert the clustered column chart visual into a scatter chart.

2. Select the chart visual, and then select **Scatter chart** on the **Visualizations** pane.

Notice the changes to the **Visualizations** pane. The **District** field is now listed under **Values**. The chart axes are also different. Make sure that Power BI plots the **Sales Per Sq Ft** field along the **X Axis** and the **Total Sales Variance %** field along the **Y Axis**.



Values
District

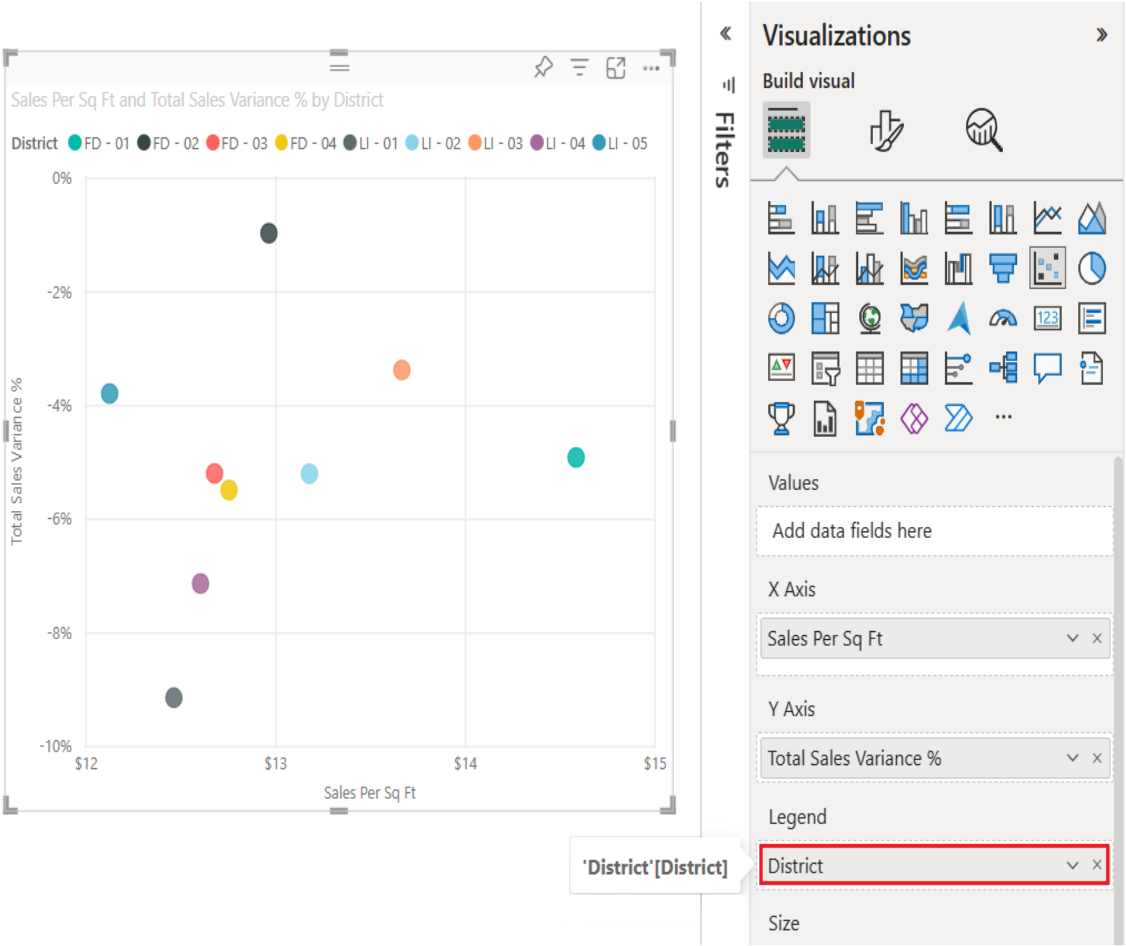
X Axis
Sales Per Sq Ft

Y Axis
Total Sales Variance %

Legend
Add data fields here

3. On the **Visualizations** pane, drag the **District** field from the **Values** section to the **Legend** section.

   Power BI creates data points where the data values intersect along the x and y axes. The data point colors represent different districts.


Sales Per Sq Ft and Total Sales Variance % by District
District ● FD - 01 ● FD - 02 ● FD - 03 ● FD - 04 ● LI - 01 ● LI - 02 ● LI - 03 ● LI - 04 ● LI - 05
0%
-2%
-4%
-6%
-8%
-10%
Total Sales Variance %
$12
$13
$14
$15
Sales Per Sq Ft
Filters
Build visual
...

Visualizations

Values

Add data fields here

X Axis

Sales Per Sq Ft

Y Axis

Total Sales Variance %

Legend

'District'[District]

District

Size

Practical-10

Aim: Given a case study: Interactive Data Analytics with Power BI Dashboard

1. HEATHROW

- Heathrow airport is an international airport in London. It is the second busiest international airport in the world after Dubai international airport. And, also the seventh-largest in terms of total passenger traffic.

THE CHALLENGE

- Being the world’s seventh busiest airport in overall passenger traffic, one can only imagine the level of efficiency and efforts expected from the airport’s ground management to keep the airport functioning properly. Managing over 2,00,000 passengers every day can be quite a challenging task for airport authorities and ground staff. Every department needs to be in absolute coordination and sync to be able to manage the passenger traffic and give them a smooth experience at the airport. At such busy airports, every day brings new challenges and uncertainties with it. Unexpected disruptions in the smooth workflow of operations at the airport disturb the entire functioning. Issues can arise due to stormy weather, delayed flights, canceled flights, shifts in jet streams, etc. disturbing the airport’s smooth functioning. Such problems send the passengers as well as airport employees into turmoil.
- The airport needed a central digitalized management system as a solution to this problem. Such a system would use the large amounts of data being produced by operational systems at the airport and transform it into useful visual insights. The interpretations produced by the BI tool can be used by airport staff for better functioning and passenger management.

THE CHANGE

- Heathrow group went with Microsoft Power BI as their business intelligence software and Microsoft Azure for cloud services. The airport has deployed Microsoft Azure technology to collect data from back-end operational systems at the airport. These systems are *check-in counters, baggage tracking systems, flight schedules, and weather tracking systems, cargo tracking* and many more.
- The operational data from these systems are forwarded to business intelligence platforms like Power BI. In Power BI, users shape this data into useful information that the airport staff can use.

- Power BI transforms the crude information into informative visuals showing different statuses and statistics of the airport systems. Then, the ground staff like baggage handlers, gate agents, air traffic controllers, etc. use this information to properly operate and manage passengers.

- Services such as Azure Stream Analytics, Azure Data Lake Analytics, and Azure SQL Database are used to *extract, clean and prepare operational data* in real-time. Thisdata is about *flight movements, security queues, passenger transfers, and immigration queues.* Ultimately, Power BI uses data from these Azure services for analysis and interpretation.
- Operational data from different data sources come into Power BI. Then Power BI tools are used to transform that data into meaningful insights with the help of visual reports, graphics, and dashboards. About 75,000 airport employees have information on their fingertips by the virtue of Power BI.

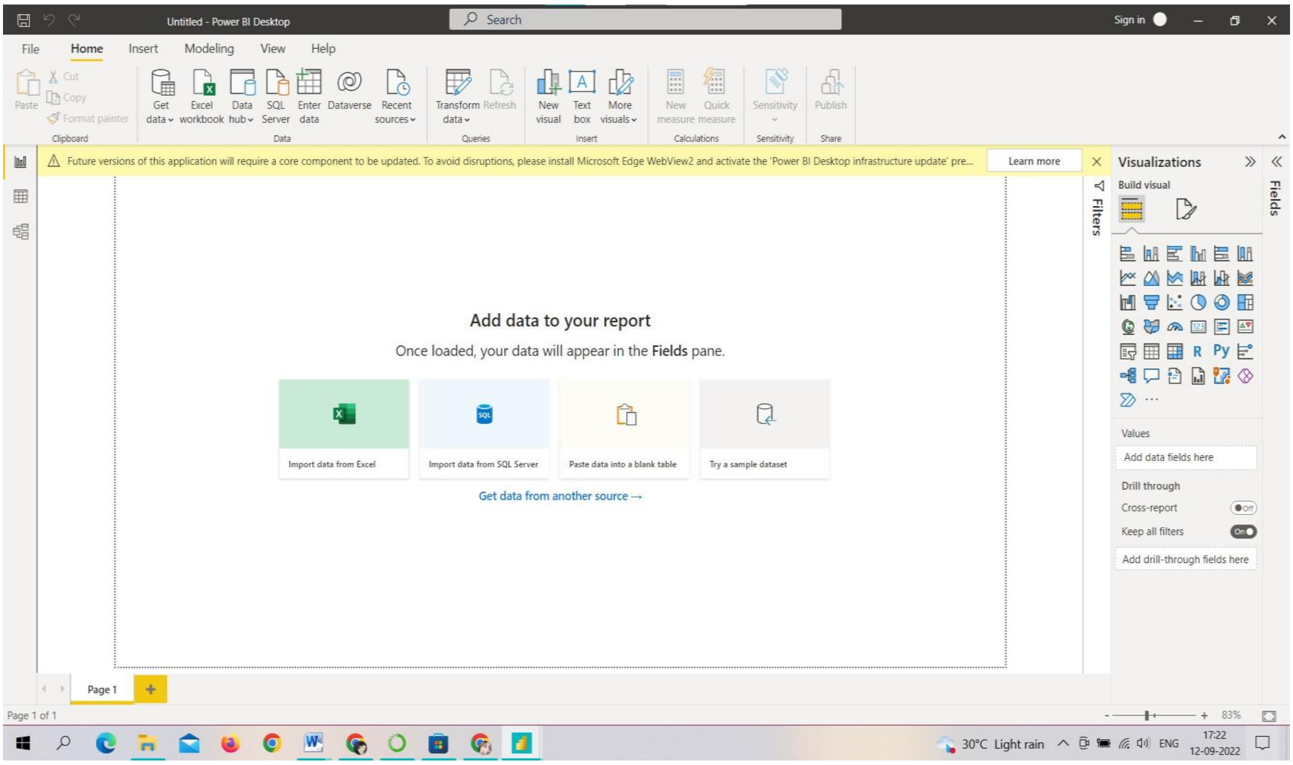
- Let us understand this with the help of a **real-world example**. If there is a change in the jet stream, it may delay about 20 flights in a day. This will result in about 6,000 passengers waiting at the airport at a given point of time. It will increase passenger traffic and density at the airport. Power BI works like the centralized information system. The airport uses it to inform about the sudden passenger influx. This information goes out to different sections such as *food outlets, immigration, customs, gate attenders, baggage handlers* at the airport. This will give them time to prepare themselves to attend the passengers.
- With the presence of smart BI solutions like Power BI, airport staff is notified in advance about the probable delays and the sudden rush of passengers at the airport. This help management groups and other employees to take suitable actions in advance like *increasing the food stock, adding extra passenger buses, increasing the ground staff, directing the passengers to the waiting area,* etc. to avoid any last-minute hustle.
- Thus, with the help of a powerful BI tool like Power BI, Heathrow has been benefited in more than one way. They are extremely happy and satisfied with the capabilities of Power BI helping them give a hassle-free airport experience to their passengers. Heathrow also is extending Power BI applications by trying to anticipate passenger flow at the airport to avoid any unexpected disruptions for the passengers.

STEPS TO ANALYZE DATA IN POWER BI

Step1: Download the Power BI Software in your PC

Step2: After that download the dataset for performing data analysis. Here I have downloaded Sales of chocolates to importing to different countries and containing info about the dealer

Step 3: Then open Power BI. The home page will look like below page



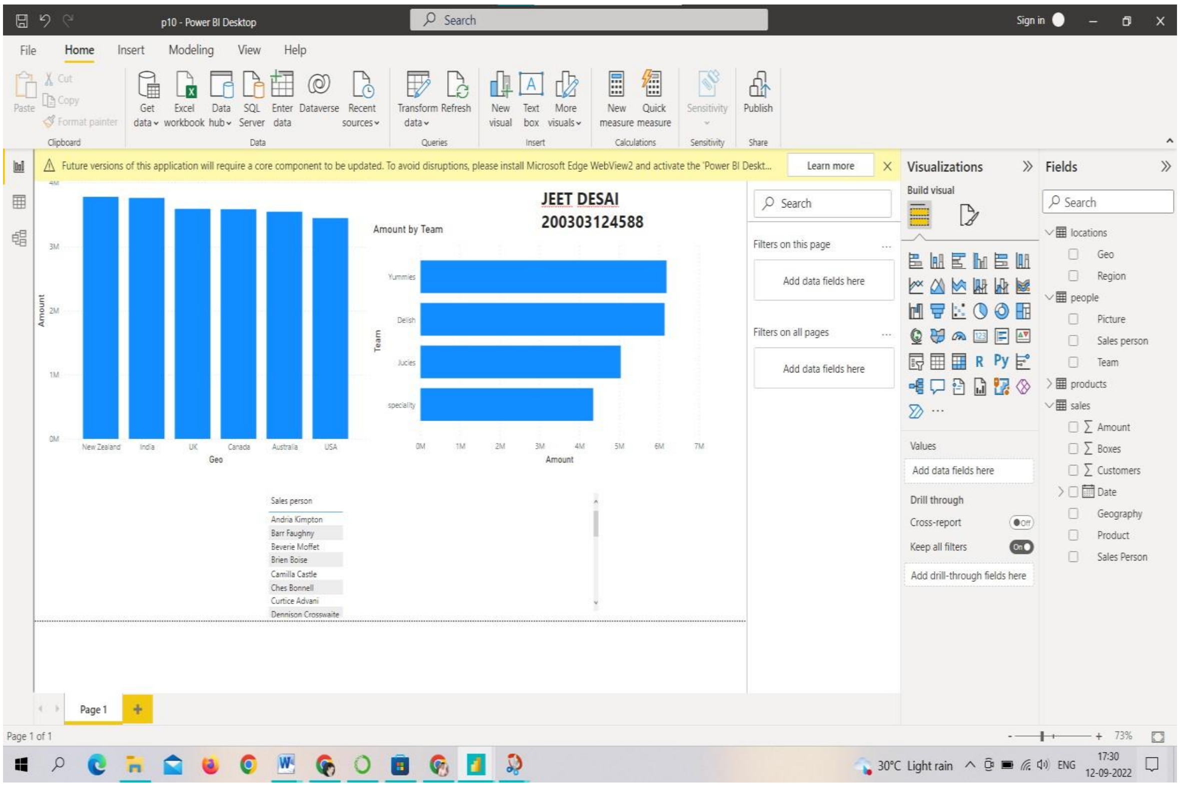
Future versions of this application will require a core component to be updated. To avoid disruptions, please install Microsoft Edge WebView2 and activate the 'Power BI Desktop infrastructure update' pre...

Add data to your report

Once loaded, your data will appear in the **Fields** pane.

Step 4: Then import file which you want

Step 5: Then according to you do data analysis



Future versions of this application will require a core component to be updated. To avoid disruptions, please install Microsoft Edge WebView2 and activate the 'Power BI Deskt...

JEET DESAI
200303124588

Sales person
Andria Kimpton
Barr Faughny
Beverie Moffet
Brien Boise
Camilla Castle
Ches Bonnell
Curtice Advani
Dennison Crosswaite