Himanshu Kumar

**Python Assignment-2 (SQL, EDA,DATA VISUALIZATION)**

For Q1 to Q3, refer to the below table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| EMPLOYEE\_ID | FIRST\_NAME | LAST\_NAME | HIRE\_DATE | JOB\_ID | SALARY |
| 100 | Steven | King | 2003-06-17 | AD\_PRES | 24000 |
| 101 | Neena | Kochhar | 2005-09-21 | AD\_VP | 17000 |
| 102 | Lex | De Haan | 2001-01-13 | AD\_VP | 17000 |
| 103 | Alexander | Hunold | 2006-01-03 | IT\_PROG | 9000 |
| 104 | Bruce | Ernst | 2007-05-21 | IT\_PROG | 6000 |
| 105 | David | Austin | 2005-06-25 | IT\_PROG | 4800 |
| 106 | Valli | Pataballa | 2006-02-05 | IT\_PROG | 4800 |
| 107 | Diana | Lorentz | 2007-02-07 | IT\_PROG | 4200 |
| 108 | Nancy | Greenberg | 2002-08-17 | FI\_MGR | 12008 |

**Q1. Construct a query to fetch all the records whose First\_name has either ‘a’ or ‘ i ‘.**

# ANS:- USE DATABASE EMPLOYEES;

SELECT \*

# FROM EMPLOYEES

where FIRST\_NAME Like ‘%a%’ OR FIRST\_NAME Like ‘%i%’ ;

**Q2. Construct a query to return Employee\_ID, First\_name, Job\_ID and salary for those employees who are either AD\_VP or AD\_Pres or their salary is greater than 10,000.**

# ANS:-

SELECT EMPLOYEE\_ID, FIRST\_NAME, JOB\_ID, SALARY FROM EMPLOYEES

# WHERE JOB\_ID IN (‘AD\_VP’ , ‘AD\_PRES’)

AND SALARY > 10000;

**Q3. Construct a query to return those records whose hire\_date is greater than 1st Jan 2005 and their last\_name doesn’t contain “a” and then sort the salary in the ascending order.**

# ANS:-

SELECT HIRE\_DATE, LAST\_NAME, SALARY FROM EMPLOYEES

# WHERE HIRE\_DATE >’2005-01-01’ AND LAST\_NAME NOT LIKE ’%a%’ ORDER BY SALARY asc;

**For Q4 and Q5, refer to the below table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ord\_no | purch\_amt | ord\_date | customer\_id | salesman\_id |
| 70001 | 150.5 | 05-10-2012 | 3005 | 5002 |
| 70009 | 270.65 | 10-09-2012 | 3001 | 5005 |
| 70002 | 65.26 | 05-10-2012 | 3002 | 5001 |
| 70004 | 110.5 | 17-08-2012 | 3009 | 5003 |
| 70007 | 948.5 | 10-09-2012 | 3005 | 5002 |
| 70005 | 2400.6 | 27-07-2012 | 3007 | 5001 |
| 70008 | 5760 | 10-09-2012 | 3002 | 5001 |
| 70010 | 1983.43 | 10-10-2012 | 3004 | 5006 |
| 70003 | 2480.4 | 10-10-2012 | 3009 | 5003 |
| 70012 | 250.45 | 27-06-2012 | 3008 | 5002 |
| 70011 | 75.29 | 17-08-2012 | 3003 | 5007 |
| 70013 | 3045.6 | 25-04-2012 | 3002 | 5001 |

**Q4. Write a SQL query to count all the orders generated on 10th Oct 2012. Return number of orders.**

# ANS:-

SELECT count(ord\_no)

# FROM orders

WHERE ord\_date= ‘10-10-2012’;

**Q5. Write a SQL query to determine the maximum order (purchase\_amt) generated by each salesman.**

# ANS:-

SELECT salesman\_id , MAX(purch\_amt) FROM orders

# GROUP BY salesman\_id;

**Q6. Write a Pandas program to select first 2 rows and first 2 columns from a dataset.**

# ANS:-

Dataset name= my\_data Import pandas as pd

# Df=pd.read\_excel(“file\_name.xlsx”) Df.iloc[0:2,0:2]

Df

**Q7. Import the iris dataset from seaborn library. Write a Python program to create a Bar plot to get the frequency of the three species of the Iris data.**

# ANS:-

df=sns.load\_dataset('iris') print(df)

# bar\_plot=sns.barplot(x="species",y="petal\_length",data=df)

* Rename the x-axis and y-axis

# ANS:-

bar\_plot=sns.barplot(x="species",y="petal\_length",data=df).set(xlabel='sepal\_width', ylabel='sepal\_length')

* Give the name to the chart title.

# ANS:-

bar\_plot=sns.barplot(x="species",y="petal\_length",data=df).set(xlabel='sepal\_width', ylabel='sepal\_length',title='Frequency of species')

**Q8. Write a Pandas program to join the two dataframes using the common column of both dataframes and return the common records.**

# ANS:-

*#Given dataset*

*Df1=”zomato.csv”*

*Df2=”country-code.xlsx”*

# Df1=pd.read\_csv(”zomato.csv”)

Df1(shift+enter)

# Df2= pd.read\_excel(”country-code.xlsx”)

Df2(shift+enter)

# (as country code is in both of the dataset then we merge both data set by ‘country code’)

Df\_final=pd.merge(Df1,Df2,on=’country code’,how=left)

**Q9. Create a dataframe of 10\*3 matrix using random function and then create a pie chart for 4th row in the dataframe. Also print the percentages on the pie chart.**

# ANS:-

import pandas as pd import seaborn as sns

import matplotlib.pyplot as plt import numpy as np

np.random.seed(100) df=pd.DataFrame(np.random.rand(10,3),index=pd.date\_range('2022-07- 06',periods=10),columns=['data','data1','data2'])

df (shift+enter)

d=df.iloc[4] d(shift+enter)

d.plot.pie(figsize=(5,5),autopct='%.2f',fontsize=10)

**Q10. Create a dataframe of 50\*23 matrix using random function. Rename the columns as New1, New2 and New3. Then create the scatterplot between New1 and new2.**

# ANS:-

import pandas as pd import seaborn as sns

# import matplotlib.pyplot as plt import numpy as np

np.random.seed(100) df=pd.DataFrame(np.random.rand(50,3),columns=['a','b','c'])

# df(shift+enter) df.rename(columns={'a':'New1','b':'New2','c':'New3'}) df(shift+enter)

df.plot.scatter(x='New1',y='New2')

**Q11. Import the Zomato dataset and create a scatter3d graph by taking 3 variables.**

# ANS:-

df=pd.read\_csv("zomato.csv") print(df)

# df.iloc[0:20].plot.scatter(c='Aggregate rating',y='Votes',x='Rating text',figsize=(12,6))

**Q12. Write a command to see the current working directory in jupyter notebook.**

# ANS:- print(pwd)

**Q13. Import the Zomato dataset and perorm the following operation:**

* **Function to see first 7 rows.**

# ANS:- df.head(7)

* Function to see last 4 rows.

# Ans:- df.tail(4)

* Function to see all the column names.

# Ans:- df.columns

**Q14. Import the titanic dataset from seaborn library and get the unique values pf Pclass.**

# Ans:-

df = sns.load\_dataset('titanic') print(df)

# df["pclass"].unique()

**Q15. In titanic dataset, perform following operations:**

* **Check the datatypes of all the variables**

# ANS:- df.dtypes

* Count the number of passengers in each Pclass category.

# ANS:-

passenger\_df=df[['pclass']].sum()

# passenger\_df=df.groupby('pclass').count() passenger\_df