================================================================================

Project - 2 --- Data Analysis With Python Pandas ----- Employee Salary DataSet

================================================================================

To perform data analysis, WE have 21 Data Analysis questions that we are going to solve one by one. Questions are as follows.

------------------------------------------

1. Display Top 10 Rows of The Dataset

2. Check Last 10 Rows of The Dataset

3. Find Shape of Our Dataset (Number of Rows And Number of Columns)

4. Getting Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And

Memory Requirement

5. Check Null Values In The Dataset

6. Drop ID, Notes, Agency, and Status Columns

7. Get Overall Statistics About The Dataframe

8. Find Occurrence of The Employee Names (Top 5)

9. Find The Number of Unique Job Titles

10.Total Number of Job Titles Contain Captain

11. Display All the Employee Names From Fire Department

12. Find Minimum, Maximum, and Average BasePay

13. Replace 'Not Provided' in EmployeeName' Column to NaN

14. Drop The Rows Having 5 Missing Values

15. Find Job Title of ALBERT PARDINI

16. How Much ALBERT PARDINI Make (Include Benefits)?

17.Display Name of The Person Having The Highest BasePay

18.Find Average BasePay of All Employee Per Year

19. Find Average BasePay of All Employee Per JobTitle

20. Find Average BasePay of Employee Having Job Title ACCOUNTANT

21. Find Top 5 Most Common Jobs

======================================================================================

Solutions for Data Analysis With Python Pandas ----- Employee Salary DataSet

======================================================================================

1. Display Top 10 Rows of The Dataset

df.head(10)

2. Check Last 10 Rows of The Dataset

df.tail(10)

3. Find Shape of Our Dataset (Number of Rows And Number of Columns)

df.shape

4. Getting Information About Our Dataset Like Total Number Rows, Total Number of Columns, Datatypes of Each Column And Memory Requirement

df.info()

5. Check Null Values In The Dataset

df.isnull().sum()

6. Drop ID, Notes, Agency, and Status Columns

df=df.drop(["id","notes","agency","status'],axis=1)

OR

df.drop(["id","notes","agency","status'],axis=1,inplace=True)

7. Get Overall Statistics About The Dataframe

df.describe() or df.describe(include="all")

8. Find Occurrence of The Employee Names (Top 5)

df["Employee Names"].value\_counts()

df["Employee Names"].value\_counts().head()

9. Find The Number of Unique Job Titles

df["Job Titles].nunique()

10.Total Number of Job Titles Contain Captain

df["Job Titles"].str.contains("CAPTAINS")---boolean array

df[df["Job Titles"].str.contains("CAPTAINS")]

len(df[df["Job Titles"].str.contains("CAPTAINS",case=False)])

OR

df[df["Job Titles"].str.contains("CAPTAINS",case=False)].count()

11. Display All the Employee Names From Fire Department

df[df["Job Titles"].str.contains("fire",case=False)]['emp name']

12. Find Minimum, Maximum, and Average BasePay

df['basicpay'].describe()

13. Replace 'Not Provided' in EmployeeName' Column to NaN

df['emp name']=df["emp name"].replace('Not Provided',np.NaN)

14. Drop The Rows Having 5 Missing Values

df.isnull().sum(axis=1)==5---boolean array

df[df.isnull().sum(axis=1)==5]--we get data with null values

df.drop(df[df.isnull().sum(axis=1)==5],axis=0)---we get Key errr

df.drop(df[df.isnull().sum(axis=1)==5].index,axis=0) #droped rows

df.drop(df[df.isnull().sum(axis=1)==5].index,axis=0,inplace=True)# for permanent

15. Find Job Title of ALBERT PARDINI

df[df["emp name"]=="ALBERT PARDINI"]["job title"]

16. How Much ALBERT PARDINI Make (Include Benefits)?

df[df["emp name"]=="ALBERT PARDINI"]["totalpaybenifits"]

17.Display Name of The Person Having The Highest BasePay

df['basicpay'].max()

df['basicpay'].max()==df['basicpay']

df[df['basicpay'].max()==df['basicpay']]['emp name']

18.Find Average BasePay of All Employee Per Year

df.groupby('year').mean()['basicpay']

19. Find Average BasePay of All Employee Per JobTitle

df.groupby('JobTtitle').mean()['basicpay']

20. Find Average BasePay of Employee Having Job Title ACCOUNTANT

df['jobtitle']=="ACCOUNTANT"

df[df['jobtitle']=="ACCOUNTANT"]['basicpay'].mean()

21. Find Top 5 Most Common Jobs

df['jobtitle'].value\_counts().head()

=================================================================================================