CoderSchool

M3.1 Python 101

Exam Solution

Part 1: MULTIPLE CHOICES

Question 1: What is the shape of the dataframe?

D. (148654, 10)

```
df.shape
```

Question 2: Assume that every employee has a different name, how many employees are there in the dataset?

A. 110811

```
df['EmployeeName'].nunique()
```

Question 3: What is the average Base Pay of the whole dataset?

A. 66325.45

```
df['BasePay'].mean()
```

Question 4: Which of the observations is FALSE about the San Francisco employees?

B. However, the average of total pay increased steadily throughout the years thanks to better payment policy.

```
df.groupby('Year').agg({'EmployeeName':'nunique',
                        'TotalPay':'mean',
                        'OvertimePay':'mean',
                        'BasePay':'max'})
```

Question 5: What is the name of the highest paid person (including benefits)?

B. NATHANIEL FORD

```
df[df['TotalPay']==df['TotalPay'].max()]['EmployeeName'].values[0]
```

Question 6: What is his job title?

A. General Manager-Metropolitan Transit Authority

```
df[df['TotalPay']==df['TotalPay'].max()]['JobTitle'].values[0]
```

Question 7: What is the name of the lowest paid person (including benefits)?

C. JOE LOPEZ

```
df[df['TotalPay']==df['TotalPay'].min()]['EmployeeName'].values[0]
```

Question 8: How much did the person in Question 7 get paid (TotalPay) in 2014?

A. -618.13

```
df[df['TotalPay']==df['TotalPay'].min()]['TotalPay'].values[0]
```

Question 9: How many unique employees have the same Job Title with Liller Jackson in the dataset?

D. 75

```
df.loc[df['JobTitle']==df.loc[df['EmployeeName']=='Liller Jackson',
'JobTitle'].values[0],'EmployeeName'].nunique()
```

Question 10: How many percent of Special Nurse get paid overtime?

C. About 26%

```
(df.loc[df['JobTitle']=='Special Nurse', 'OvertimePay']>0).mean()
```

Question 11: Among employees who have TotalPay higher than the average TotalPay, how many unique Job Titles do they hold?

C. 1307

```
df.loc[df['TotalPay']>df['TotalPay'].mean(), 'JobTitle'].nunique()
```

Question 12: How many Job Titles were represented by only one person in 2013?

A. 202

```
(df.loc[df['Year'] == 2013, 'JobTitle'].value_counts() == 1).sum()
```

Part 2: EXPRESSIONS

Question 13: Write a single line of code to return all employees whose TotalPay is higher than the average TotalPay.

```
df[df['TotalPay']>df['TotalPay'].mean()]
```

Question 14: Write a single line of code to return the top 5 most common jobs.

```
df['JobTitle'].value_counts().head()
```

Question 15: Among those 5 most common Job Titles of the whole dataset, how have the average Base Pay, Overtime Pay, and Total Pay changed over 4 years of the dataset?

Hint: Start by identifying the most common jobs, then filter the dataset before computing the summary.

```
pd.pivot_table(data=df[df['JobTitle'].isin(df['JobTitle'].value_counts(
).head().index)],
               index=['JobTitle'],
               columns=['Year'],
               values=['BasePay', 'OvertimePay', 'TotalPay'])
```