Assignment #1 Report

Healthcare dataset

1-What is the total number of features (independent variables) in the dataset, and what are their names?

As I understand; The number of features is different from the number of the independent variables.
The number of features includes all columns : 15.
Number of independent variables : Not determined yet.

```
#number of features
len(dataset.columns)
```

```
15
```

2-Which blood types are represented in the dataset?

['B-', 'A+', 'A-', 'O+', 'AB+', 'AB-', 'B+', 'O-']

```
#blood types
dataset['Blood Type'].unique()
```

```
array(['B-', 'A+', 'A-', 'O+', 'AB+', 'AB-', 'B+', 'O-'], dtype=object)
```

3-Based on the dataset, which blood type is the most common, and which is the rarest?

Most frequent : A-
Rarest : O-

```
[20] #most frequent blood type
     dataset['Blood Type'].mode() #or: dataset['Blood Type'].value_counts().idxmax()
```

	Blood Type
0	A-

dtype: object

```
#rarest blood type
dataset['Blood Type'].value_counts().idxmin()
```

```
'O-'
```

4-How many different medical conditions are included in the dataset?

6

```
#medical conditions
len(dataset['Medical Condition'].unique())
```

```
6
```

5-Who is the youngest and oldest patient in the dataset?

Youngest: jamES BasS phD
Oldest: DAVId NeWTOn

```
[40] #youngest patient
     youngest = dataset[dataset['Age'] == dataset['Age'].min()]
     print('The name of the youngest patient is: ', youngest['Name'].iloc[0],
           '\nand his age is ', dataset['Age'].min())
```

```
The name of the youngest patient is:  jamES BasS phD
and his age is  13
```

```
#oldest patient
youngest = dataset[dataset['Age'] == dataset['Age'].max()]
print('The name of the oldest patient is: ', youngest['Name'].iloc[0],
      '\nand his age is ', dataset['Age'].max())
```

```
The name of the oldest patient is:  DAVId NeWTOn
and his age is  89
```

6-What is the average billing amount, and what is the total billing amount across all patients?

Average billing amount: 25539.316097211795
Total billing amount: 1417432043.3952546

```
[42] #average billing amount
     dataset['Billing Amount'].mean()
```

```
25539.316097211795
```

```
#total billing amount across all patients
dataset['Billing Amount'].sum()
```

```
1417432043.3952546
```

Salary dataset

1.Select all columns between the 2nd and 8th columns, and rows from 1 to 1000.

dataset2.iloc[1:1001, 1:9] #note: I included the 2nd and 8th columns(from column 2 to 8)

```
#columns between the 2nd and 8th columns, and rows from 1 to 1000
dataset2.iloc[1:1001, 1:9] #note: I included the 2nd and 8th cloumns(from column 2 to 8)
```

	EmployeeName	JobTitle	BasePay	OvertimePay	OtherPay	Benefits	TotalPay	TotalPayBenefits
1	GARY JIMENEZ	CAPTAIN III (POLICE DEPARTMENT)	155966.02	245131.88	137811.38	NaN	538909.28	538909.28
2	ALBERT PARDINI	CAPTAIN III (POLICE DEPARTMENT)	212739.13	106088.18	16452.60	NaN	335279.91	335279.91
3	CHRISTOPHER CHONG	WIRE ROPE CABLE MAINTENANCE MECHANIC	77916.00	56120.71	198306.90	NaN	332343.61	332343.61
4	PATRICK GARDNER	DEPUTY CHIEF OF DEPARTMENT,(FIRE DEPARTMENT)	134401.60	9737.00	182234.59	NaN	326373.19	326373.19
5	DAVID SULLIVAN	ASSISTANT DEPUTY CHIEF II	118602.00	8601.00	189082.74	NaN	316285.74	316285.74
...	...	...	...	...	...	...	...	...
996	KATHERINE MAH	CLINICAL NURSE SPECIALIST	169572.06	0.00	250.00	NaN	169822.06	169822.06
997	CHRISTINE DAVENPORT	CLINICAL NURSE SPECIALIST	169572.05	0.00	250.00	NaN	169822.05	169822.05
998	ANNE HUGHES	CLINICAL NURSE SPECIALIST	169572.01	0.00	250.00	NaN	169822.01	169822.01
999	JOHN FAZIO	CLINICAL NURSE SPECIALIST	169572.00	0.00	250.00	NaN	169822.00	169822.00
1000	ROBERT KUZMA	CAPTAIN, EMERGENCYCY MEDICAL SERVICES	140546.84	8577.69	20692.86	NaN	169817.39	169817.39

1000 rows × 8 columns

2.How many missing values are there for each feature, and what is the total count of missing values in the dataset?

Number of missing values are there for each feature:

```
#number of missing values for each feature
dataset2.isnull().sum()
```

	0
Id	0
EmployeeName	0
JobTitle	0
BasePay	609
OvertimePay	4
OtherPay	4
Benefits	36163
TotalPay	0
TotalPayBenefits	0
Year	0
Notes	148654
Agency	0
Status	148654

dtype: int64

Total count of missing values in the dataset: 334088

```
#total count of the missing values in the dataset
dataset2.isnull().sum().sum()
```

```
334088
```

3. What is the median value for the "basepay" feature?

Number of missing values are there for each feature: 65007.45

```
#median value for the "basepay" feature
dataset2['BasePay'].median()
```

```
65007.45
```

Done by: Mahar Zeyad.