Tasks Solutions

View Data:

Task1 = #Show the first seven rows in the dataset

```
print("\nFirst seven rows of Data Sest:")
display(df.head(7))
```

Task2 = #Show the last five rows in the dataset

```
print("\nLast five rows of Data Sest:")
display(df.tail())
```

Task3 = #Show number of rows

```
print("\nNumber of Rows: ")
df.shape[0]
```

Task4 = #Show number of columns

```
print("\nNumber of Coulmns: ")
df.shape[1]
```

Task5 = # Get categorical column names

```
categorical_cols = df.select_dtypes(include=['object']).columns
```

Task6 = # Get numerical column names

```
num_df = df.select_dtypes(include=['int64', 'float64']).columns
```

Task7 = # show sub-dataset with only categorical columns

```
display(df.select_dtypes(include=[‘object’])
```

Task8 = # show sub-dataset with only numerical columns

```
display(df.select_dtypes(include=[‘int64’,’float64’])
```

Task9 = #Show the null values

```
print("\nSum of null values:")
df.isnull().sum()
```

Slicing and indexing

Tasks - do all what we have done but using the .iloc method instead

df.iloc[:,6]

df.iloc[:,14] or df.iloc[:,-1]

df.iloc[:,[3,4,5]]

df.iloc[:,[9,10,11]]

df.iloc[:,7:]

df.iloc[:,12:15]

Cleaning and Handling Missing Values Tasks

Task 1 - Check total numbers of null values in each column

```
df.isnull().sum()
```

Task 2 - Replace all unstandard missing values in the remaining columns

```
df['col_name'] = df['col_name'].replace('?', np.nan)
```

Task 3 - handle the remaining of the missing values with filling them in, imputing them, and drop missing values from absences column.

```
df['col_name'] = df['col_name'].fillna('specific value')
df['col_name'] = df[col_name'].fillna(df['col_name'].mode()[0])
df['col_name'] = df[col_name'].fillna(df['col_name'].mean())
df['col_name'] = df[col_name'].fillna(df['col_name'].median())
df['col_name'].dropna(inplace=True)
```