Name: Krushnakumar Patle

Email: krishnapatle128@gmail.com

Batch: Data Engineering Batch-1

Python Coding Challenge

Q1. Explain Pandas for Data Processing & execute Reading CSV Data using Pandas &Read Data from CSV Files to Pandas Dataframes &Filter Data in Pandas Dataframe using query.

Pandas is a powerful library for data processing in Python. It provides two main data structures: Series (one-dimensional) and DataFrame (two-dimensional), which are highly efficient for working with structured data. Pandas simplifies many common data processing tasks, including reading and writing data in various formats, cleaning and transforming data, handling missing values, and performing complex operations on datasets.

```
# I have uploaded the csv here
from google.colab import files
uploaded = files.upload()
```

```
Choose Files student-dataset.csv
• student-dataset.csv(text/csv) - 27433 bytes, last modified: 1/30/2024 - 100% done
Saving student-dataset.csv to student-dataset.csv
```

```
[2] import io
    # Check if the uploaded dictionary contains the file
    if 'student-dataset.csv' in uploaded:
        data = io.BytesIO(uploaded['student-dataset.csv'])
    else:
        print("The uploaded dictionary doesn't contain the 'student-dataset.csv' file.")
```

Execute Reading CSV Data using Pandas

```
import pandas as pd

# reading csv file
df = pd.read_csv("student-dataset.csv")
print(df.head())
```

```
   id             name               nationality                 city  \
0   0        Kiana Lor                     China              Suzhou
1   1   Joshua Lonaker  United States of America       Santa Clarita
2   2    Dakota Blanco  United States of America             Oakland
3   3  Natasha Yarusso  United States of America       Castro Valley
4   4   Brooke Cazares                    Brazil  São José dos Campos

   latitude  longitude gender  ethnic.group  age  english.grade  math.grade  \
0     31.31     120.62      F           NaN   22            3.5         3.7
1     34.39    -118.54      M           NaN   22            2.9         3.2
2     37.80    -122.27      F           NaN   22            3.9         3.8
3     37.69    -122.09      F           NaN   20            3.3         2.8
4    -23.18     -45.88      F           NaN   21            3.7         2.6

   sciences.grade  language.grade  portfolio.rating  coverletter.rating  \
0             3.1             1.0                 4                 4.0
1             3.6             5.0                 5                 4.0
2             3.2             5.0                 3                 3.0
3             3.2             5.0                 5                 2.0
4             3.4             1.0                 4                 4.0

   refletter.rating
0                 4
1                 5
2                 4
```

```
[3]  1    5
     2    4
     3    4
     4    5
```

Read Data from CSV Files to Pandas Dataframes

```
df = pd.read_csv('student-dataset.csv',
        header=0,
        usecols=["name", "nationality", "city"])
# printing dataframe
print(df.head())
```

```
              name              nationality                city
0       Kiana Lor                    China             Suzhou
1   Joshua Lonaker  United States of America       Santa Clarita
2    Dakota Blanco  United States of America             Oakland
3  Natasha Yarusso  United States of America       Castro Valley
4   Brooke Cazares                   Brazil  São José dos Campos
```

Filter Data in Pandas Dataframe using query.

```
[5]  filtered_data=df[df['name']=="Joshua Lonaker"]
     print(filtered_data)
```

```
              name              nationality          city
1  Joshua Lonaker  United States of America  Santa Clarita
```

Q2. Execute with one example Lambda Functions in Python&Read JSON Strings to Python dicts or lists

Lambda FUnction:

Python Lambda Functions are anonymous functions means that the function is without a name. As we already know the def keyword is used to define a normal function in Python. Similarly, the lambda keyword is used to define an anonymous function in Python.

JSON:

JSON (JavaScript Object Notation) is a lightweight data interchange format that is easy for humans to read and write, and easy for machines to parse and generate. It is a text format that is language-independent and is commonly used to transmit data between a server and a web application as an alternative to XML.

```
[6]  #this is program for finding square by using lambda function
     square = lambda x: x ** 2
     result = square(4)
     print(result)
```

```
16
```

Read JSON Strings to Python dicts or lists

```python
import json

# JSON string
jsonString = '{ "id": 121, "name": "Krishna", "course": "Data Engineering"}'

# It Convert JSON String to Python
student_details = json.loads(jsonString)

# It Print Dictionary
print(student_details)

print(student_details['name'])
print(student_details['course'])
```

```
{'id': 121, 'name': 'Krishna', 'course': 'Data Engineering'}
Krishna
Data Engineering
```

```python
# this read JSON string to list in Pandas

import pandas as pd

# JSON string
json_data = '{"Name": ["Krishna", "Ajay", "Jayant"], "Age": [22, 23, 25]}'

# Read JSON string into a Pandas DataFrame
df_json = pd.read_json(json_data)
```

```python
# Read JSON string into a Pandas DataFrame
df_json = pd.read_json(json_data)

# Display the DataFrame
print(df_json)
```

```
      Name  Age
0  Krishna   22
1     Ajay   23
2   Jayant   25
```