

# Assignment NonAcademy

Sabit Hasan Student ID: NA-0F1483

## Part A

### Task 1

ডেটা অ্যানালিটিক্স কী?

ডেটা অ্যানালিটিক্স হলো তথ্য সংগ্রহ, বিশ্লেষণ এবং ব্যাখ্যার একটি প্রক্রিয়া, যার মাধ্যমে তথ্য থেকে মূল্যবান সিদ্ধান্ত নেওয়া যায়।

ইতিহাসে গুরুত্বপূর্ণ ২টি ঘটনা।

**1960s** - ডেটাবেস ম্যানেজমেন্ট সিস্টেম (**DBMS**) চালু হয়। এতে ডেটা সংরক্ষণ সহজ হয়।

**2010**-এর পর বিগ ডেটা ও ক্লাউড প্রযুক্তি ডেটা অ্যানালিটিক্সকে দ্রুত ও স্কেলেবল করে তোলে

বর্তমানের ৩টি ট্রেন্ড।

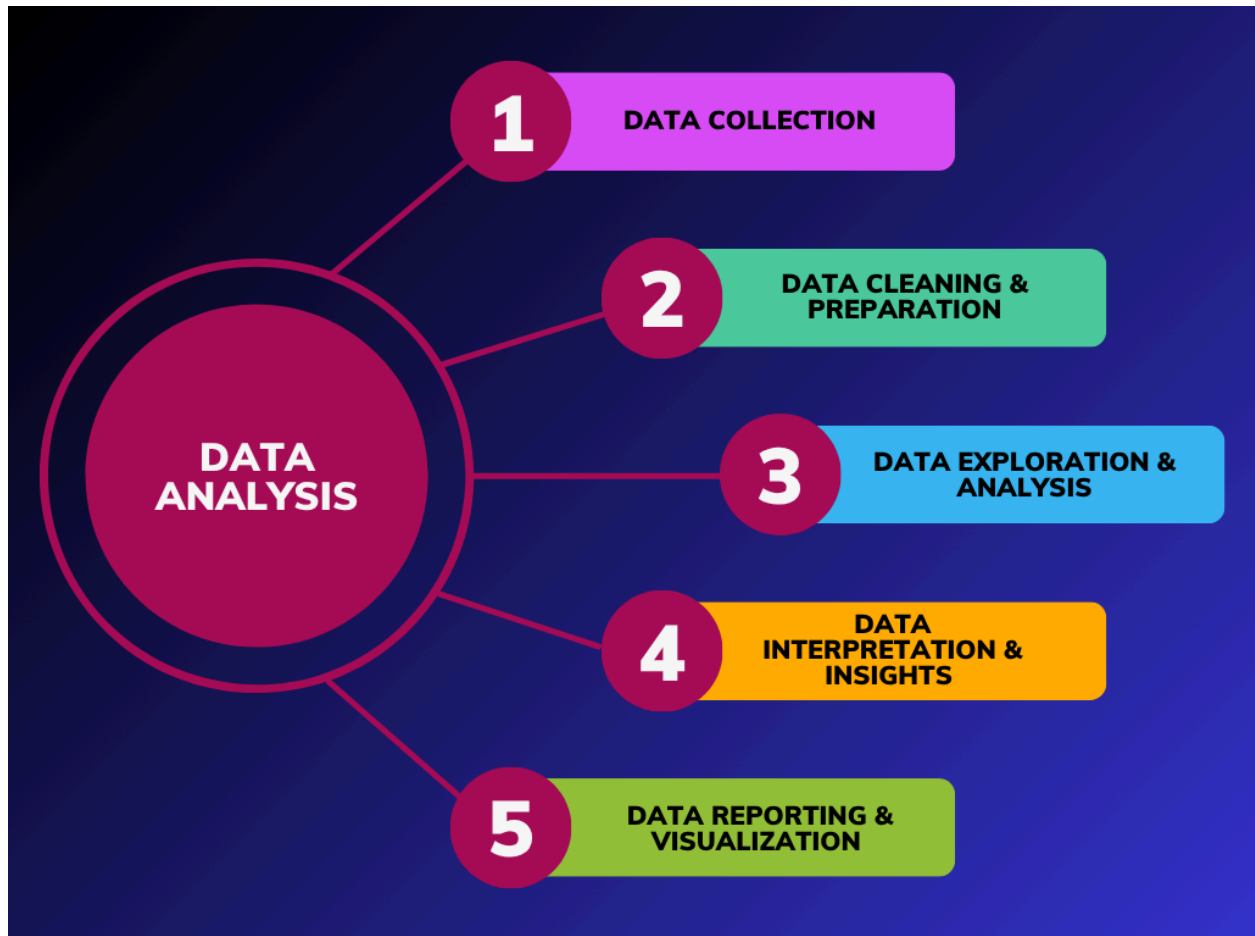
**AI ও ML**

**Cloud Analytics**

**Real-time Analytics**

# Task 2

## Analytics Process Mapping



Img: Canva

## Task 3

Term	Definition
Data Types	ডেটার ধরণ (int, float, string)
Variables	কোন তথ্য ধারণ করে এমন উপাদান
Metrics	পরিমাপযোগ্য মান যেমন: বিক্রয় পরিমাণ
KPIs	কোনো ব্যবসার সাফল্য নির্দেশ করে।
Structured Data	সারিবদ্ধ ডেটা যেমন Excel টেবিল
API	এক প্রোগ্রাম থেকে অন্যটিতে ডেটা আদান-প্রদান
SQL	Structured Query Language
Data Warehouse	বিশাল ডেটা সংরক্ষণের স্থান
Pandas	পাইথনের ডেটা অ্যানালাইসিস লাইব্রেরি
Tidyverse	আর প্রোগ্রামিংয়ের ডেটা টুলসের প্যাকেজ

# Part B

## Task 4

Source: <https://www.kaggle.com/datasets/barisk7/uber-dataset?resource=download>

Format of the data: .csv

Data Structure: structured

Ethical Considerations: কোনো ব্যক্তিগত নাম বা ঠিকানা নেই |

# Task 5

```
import pandas as pd
import seaborn as see
import matplotlib.pyplot as plt

ab = pd.read_csv(r"D:\assigenment\UberDataset.csv")

# Summary
print(ab.info())

print('-----')
#Just for margin

print(ab.describe())

print('-----')
#Just for margin

# Null values
print(ab.isnull().sum())

print('-----')
#Just for margin

# Duplicates
print(ab.duplicated())
print(ab.duplicated().sum())

print('-----')
#Just for margin

# Outliers
aa = see.boxplot(data=ab['MILES'])
print(aa)
```

## Task 6

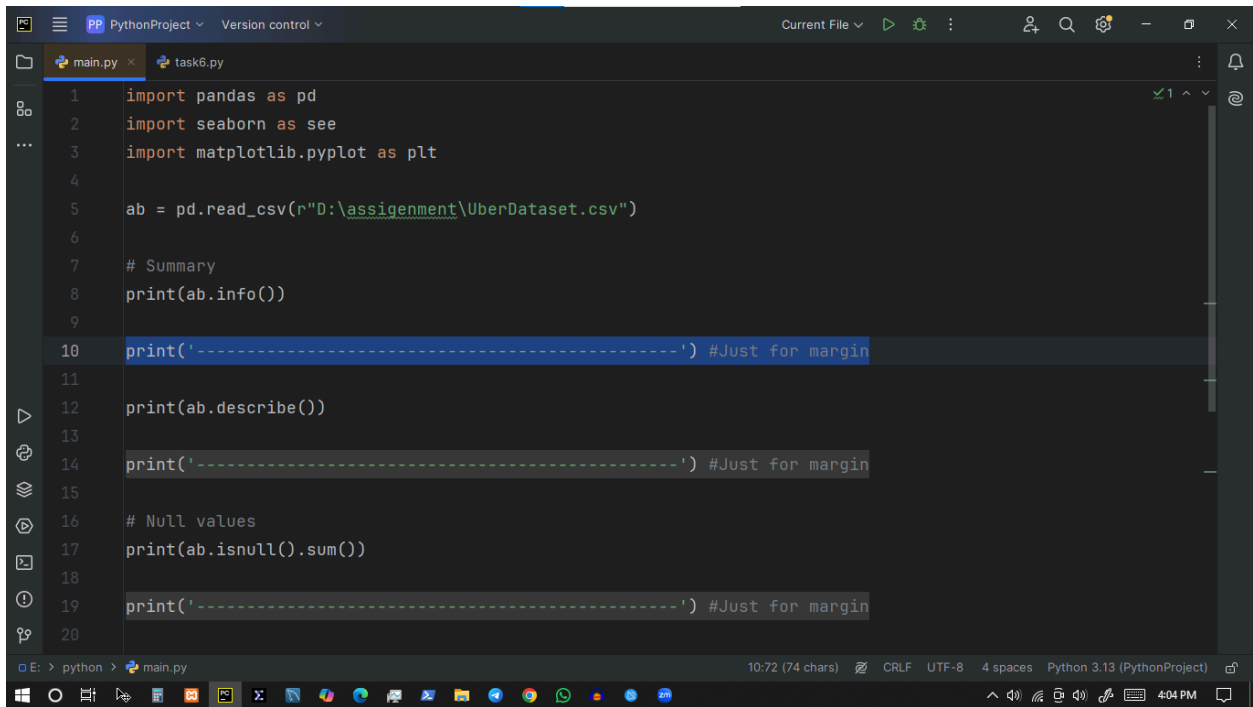
```
import pandas as pd
import seaborn as see
import matplotlib.pyplot as plt

ab = pd.read_csv(r"D:\assigenment\UberDataset.csv")

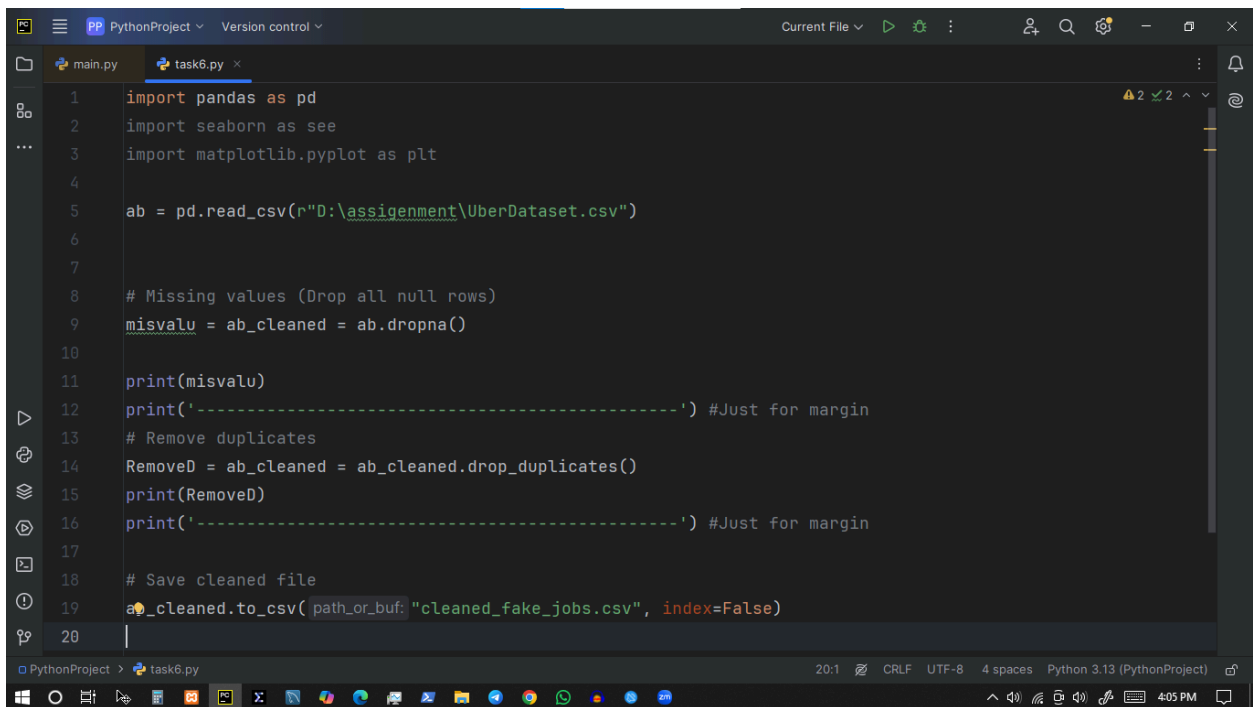
# Missing values (Drop all null rows)
misvalu = ab_cleaned = ab.dropna()

print(misvalu)
print('-----')
#Just for margin
# Remove duplicates
RemovedD = ab_cleaned = ab_cleaned.drop_duplicates()
print(RemovedD)
print('-----')
#Just for margin

# Save cleaned file
ab_cleaned.to_csv("cleaned_fake_jobs.csv", index=False)
```



```
1 import pandas as pd
2 import seaborn as see
3 import matplotlib.pyplot as plt
4
5 ab = pd.read_csv(r"D:\assignment\UberDataset.csv")
6
7 # Summary
8 print(ab.info())
9
10 print('-----') #Just for margin
11
12 print(ab.describe())
13
14 print('-----') #Just for margin
15
16 # Null values
17 print(ab.isnull().sum())
18
19 print('-----') #Just for margin
20
```



```
1 import pandas as pd
2 import seaborn as see
3 import matplotlib.pyplot as plt
4
5 ab = pd.read_csv(r"D:\assignment\UberDataset.csv")
6
7
8 # Missing values (Drop all null rows)
9 misvalu = ab_cleaned = ab.dropna()
10
11 print(misvalu)
12 print('-----') #Just for margin
13
14 # Remove duplicates
15 Removed = ab_cleaned = ab_cleaned.drop_duplicates()
16 print(Removed)
17 print('-----') #Just for margin
18
19 # Save cleaned file
20 ab_cleaned.to_csv(path_or_buf: "cleaned_fake_jobs.csv", index=False)
```

## Task 7

প্রথমত আমি এসাইনমেন্টটা ভালোভাবে বোঝার চেষ্টা করেছি, এসাইনমেন্টটা ভালোভাবে বোঝার পরে,

**Task 1** ডেটা এনালিটিক্স কি এবং বর্তমান ট্রেন্ড নিয়ে আলোচনা করা হয়েছে,

**Task 2** Canva এর মাধ্যমে ডাটা প্রসেসিং ম্যাপিং তৈরি করা হয়েছে,

**Task 3** কিছু গুরুত্বপূর্ণ Term এবং সেগুলোর ডেফিনেশন নিয়ে আলোচনা করা হয়েছে,

**Task 4** [kaggle](https://www.kaggle.com/) থেকে একটি ডাটা সেট ডাউনলোড করা হয়েছে এবং সেটার কন্ডিশন এবং স্ট্রাকচার বর্ণনা করা হয়েছে,

**Task 5** এবং **6** Python এর মাধ্যমে ডাটা কোয়ালিটি চেকিং এবং ডেটা ক্লিনিং

**Task 7** Describe our my project

**Task 8** একটি Mini Case Study about healthcare

## Task 8



## ডোমেইন: স্বাস্থ্যসেবা (Healthcare)

### ব্যখ্যা:

ডাটা অ্যানালিটিক্স-এর মাধ্যমে হাসপাতালে রোগীদের তথ্য বিশ্লেষণ করে রোগ নির্ণয় ও ওষুধের ডোজ নির্ধারণ করা যায়। আধুনিক স্বাস্থ্যসেবায় প্রতিটি রোগীর বয়স, পূর্ববর্তী চিকিৎসা ইতিহাস, ল্যাব রিপোর্ট, ওষুধের প্রতিক্রিয়া ইত্যাদি বিশ্লেষণ করে সঠিক চিকিৎসা সিদ্ধান্ত নেওয়া হয়।

উদাহরণস্বরূপ, একটি হাসপাতাল যদি ১০,০০০ রোগীর রক্তচাপ ও সুগারের রেকর্ড বিশ্লেষণ করে, তবে তারা বুঝতে পারবে কোন বয়সী রোগীদের ডায়াবেটিসের ঝুঁকি বেশি। একইভাবে, AI-ভিত্তিক ডেটা অ্যানালিটিক্স ব্যবহার করে একজন ডাক্তার রোগের সম্ভাব্য কারণ অনুমান করতে পারে এবং প্রাসঙ্গিক ওষুধ দিতে পারে।

### ব্যবহৃত ডেটার ধরন:

রোগীর ব্যক্তিগত তথ্য (বয়স, লিঙ্গ)

মেডিকেল রেকর্ড (BP, সুগার লেভেল, টেম্পারেচার)

ডায়াগনস্টিক রিপোর্ট (এক্স-রে, ব্লাড টেস্ট)

### ডেটা বিশ্লেষণের মাধ্যমে যা নির্ধারিত হয়:

রোগ নির্ণয় (Diagnosis)

ওষুধের ডোজ ও ধরন

রোগের ঝুঁকি মূল্যায়ন

চিকিৎসা ফলাফল ট্র্যাকিং