**Capstone  Project**

**CLOTHING BRANDS**

This capstone is divided into 4 Phase(weekly) duration. In this document the entire details of 4 Phase(weekly) duration is given.

**Phase 1**                                                                                                                                            21 marks

 Web scraping

A clothing brand is a business or label that operates within the fashion industry, specializing in creating and selling clothing and apparel. It encompasses designing, manufacturing, marketing, and distributing garments, targeting specific customer segments or the target market.

Here ,you have to collect Men & women /Kids wear data .

Web scraping the below given  URL

[https//www.adidas.co.in/](https://www.adidas.co.in/)

<https://oldnavy.gap.com/>

<https://www2.hm.com/en_in/index.html>

<https://allensolly.abfrl.in/>

<https://www.biba.in/>

<https://www.fabindia.com/>

<https://www2.hm.com/en_in/index.html>

<https://www.levi.in/>

<https://www.montecarlo.in/>

<https://www.muftijeans.in/>

<https://www.peterengland.com/>.

<https://uspoloassn.in/>

<https://in.benetton.com/>

<https://wforwoman.com/>.

<https://www.zodiaconline.com/>

<https://www.zara.com/in/>

<https://www.manyavar.com/>

<https://www.bottegaveneta.com/en-en>

<https://www.newbalance.com/>

<https://in.lining.studio/category/clothing>

<https://www.loewe.com/int/en/home?country=IN&gad_source=1&gclid=EAIaIQobChMI6dfnvNqbhAMVOSmDAx3DKAAAEAAYASAAEgINmvD_BwE>

<https://www.provogue.com/>

Adds –on

You are free to scrap any other new website related to clothing brand .

·         **You need to scrap minimum of 4000 entries/rows  for your dataset. The more is better.**

**Phase 2                                 (**15 SQLqueries \*1 each) + (7 join queries\*2 ) = 29 marks

Make the table in  csv format  (3 csv files)

**Table1.csv**

Sno , Brandname, Category  , Producttype   ,ProductName    ,productcode  ,price

***Details of TABLE 1***

The table, named “Table1,” contains the following columns:

1.    **Sno**: A unique identifier for each record.

2.    **Brandname**: The name of the brand associated with the product.

3.    **Category**: The category to which the product belongs -- Men/women/kids

4.    **Producttype**: The type or classification of the product.-- (Sportswear /Formal Shirts/Trouser/etc)

5.    **ProductName**: The specific name of the product.

6.    **Productcode**: A code assigned to uniquely identify the product.-- Style Code

7.    **Price**: The price of the product.

**Table2.csv**

Sno , color1,color2,color3,color4 ,Weartype , Material type ,reviews

***Details of TABLE 2***

Weartype ----topwear/bottomwear

Material type  ---- type of fabric

**Table3.csv**

Sno, Size1, Size2, Size3, Size4,  Country of origin, return time ,occasion

***Details of TABLE 3***

Size --- different sizes available of the particular product  (small/medium/large/Extra large )

Just like S /M/L /XL  etc..

Country of origin --- product manufactured in the country

Return time --- return period of the product

Occasion---(regular/casual /celebration)

**Write the SQL queries**

TABLE 1

1)      Write a query to **Retrieve the Top 20  Most Expensive Products**

2)      Write a query **Calculate the Total Price for Each Category**:

3)      Write a query to **Retrieve Products with a Price Range** 2000 to 5000

4)       Write a query to **Calculate the Average Price for Each Brand .**

5)       Write a query to **find Products with Unique Brand-Category Combinations**

Table2

1)      Write a query **Order Rows by Material Type in Ascending Order .**

2)      **Write a query to Calculate the Average Number of Reviews**

3)    **Write a query to Retrieve the Top 30 Most Expensive Products by Weartype**:

4)      Write a query to **Find Products with Similar Colors**.

5)      Write a query to **Calculate the Total Price for Each Material Type**.

Table 3 .csv

1)      Write a query to **Order Rows by Return Time in Ascending Order.**

2)      Write a query to find distinct country of origin .

3)      Write a query to **Calculate the Average Return Time by Country of Origin**.

4)      Write a query to **Find Occasions with High Return Times.**

5 ) Write a query to **Retrieve Products with Multiple Sizes (Size1, Size2, Size3, Size4)**:

**JOIN QUERY  using table1,table2 and table3**

1)      Write a query  find Brandname, Category , Producttype ,ProductName  with a **listed price greater than the average listed price** in table 1.

2) Write a query to list all Brandname, Category , Producttype ,ProductName   along with their corresponding **color information** from Table2

3)      Write a query to find the **average reviews** for each Producttype ,ProductName   using table 1 and table 2

4)     Write a query  to find products with a Product name, **material type** that matches the **most common material type** in (use table1 and table 2)

5)  Write a query   to list all products  Weartype , Material type  along with their corresponding **sizes** from   (use table 2 and table 3)

   6)

 Write a query  to find the **average return time** for each product type (use table1 and table 3)

7

Write a query to find Country of origin, return time ,occasion of each product type .

(use table 1 and table 3)

**Phase  3**30 marks

**Now make only 1 dataframe of 3 csv file using concat/merge /join operation of pandas and start doing EDA .**

**Do the complete EDA in details to explore the insights of data and write the detailed observations of each analysis .**

**Phase 4** 20marks

**Write the complete Machine learning code to make predictions of price and occasion .Use appropriate models on their label basis. Remember you need to make 2 different predictions: price and occasion   .**

**Apply all the best techniques of scaling, hyperparameter tuning, avoid underfitting or overfitting (bias/variance)**

**At the end save the best model and convey on which basis you have chosen that model.**

Submission guidelines :

1)  You need to submit  only 3 files in Github . Only Github link will be accepted .

2)  File 1 :  webscraping phase 1 file

3)  File 2 :   SQL Queries  phase 2 file

4) File 3 : EDA and Machine Learning  file  .

Note:

1)  Make one repository on github and send link on last day of submission date.

2)  Do not submit direct file in LMS. Only Github link will be accepted.

3)  No extension will be provided.

4)  Marks criteria will be followed.

5)  Plagiarism work/ Use of any AI tool  will not be accepted or cancelled.

**END**