Assignment for Rentomojo

Set(1)

Q1-

Firstly we need to clean our data as some entries are not acceptable because if we have any movie it is common to understand the runtime is always greater than 0 minutes and some data are also not available. After handling missing data and inappropriate values we perform our analysis

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import json

df = pd.read_csv("movie_dataset.csv")

df.isna().sum()//count of na values

df.fillna(0)//filling missing values

df.columns//looking at all available column name

df.shape

#as our data have NA values and Zero values

df.describe()

# rm duplicate values

sum(df.duplicated())

#print count of rows in which row have inappropriate values

print(df[(df['runtime']==0)].shape[0])

df1 = df[df.runtime != 0]

print(df1[(df1['runtime']==0)].shape[0])

#now after removng all unnecessary data and cleaning it we can use to get result

r= df1[df1['runtime'] == min(df1['runtime'])]

print("Movie with minimum runtime :"+"\n",r[['original_title','runtime']])
```

#the above statement will give output for movie with minimum runtime

```
r1= df1[df1['runtime'] == max(df1['runtime'])]
print("Movie with maximum runtime :"+"\n",r1[['original_title','runtime']])
```

#the above statement will print movie with maximum runtime

Q2

```
inf = pd.DataFrame(df1['budget'].sort_values(ascending = False))
inf['production_companies'] = df1['production_companies']
df3=inf[['production_companies','budget']]
df3=df3.head(5)
df3
l=list()
for i in df3['production_companies']:
   l.append(eval(i))
df4=pd.concat([pd.DataFrame(d) for d in l])
sum(df4.duplicated()) // find duplicate values
df4=df4.drop_duplicates()
sum(df4.duplicated())//after removing duplicate values
df4=df4.head(5)
df4//it will print top 5 production house by budget
inf = pd.DataFrame(df1['budget'].sort_values(ascending = False))
inf['production_companies'] = df1['production_companies']
df3=inf[['production_companies','budget']]
df3[:5]
for i in df4['name']:
   select_color = df1.loc[df1['production_companies'].str.contains('i')]
select_color//new dataframe with required columns
for i in df4['name']:
```

```
print("for Production house: ",i)
select_color = df1.loc[df1['production_companies'].str.contains('i')]
inf= pd.DataFrame(select_color['popularity'].sort_values(ascending=False).head())
inf['production_companies'] = select_color['production_companies']
inf['original_title'] = select_color['original_title']
inf['vote_average'] = select_color['vote_average']
inf['revenue'] = select_color['revenue']
print(inf[['original_title','revenue','vote_average']])
```

3.

_

//Now this above statement will print our final result.

4. as I am going to uses all my life savings. I will choose marvel studios because as I had done one assignment on tableau on character of marvels and dc I found that people are more into marvel heros. Although there is not much difference but Since marvel is launching new series and movies which are attracted to everyage of people .So, in future it is kinda predicted that it will grow further.

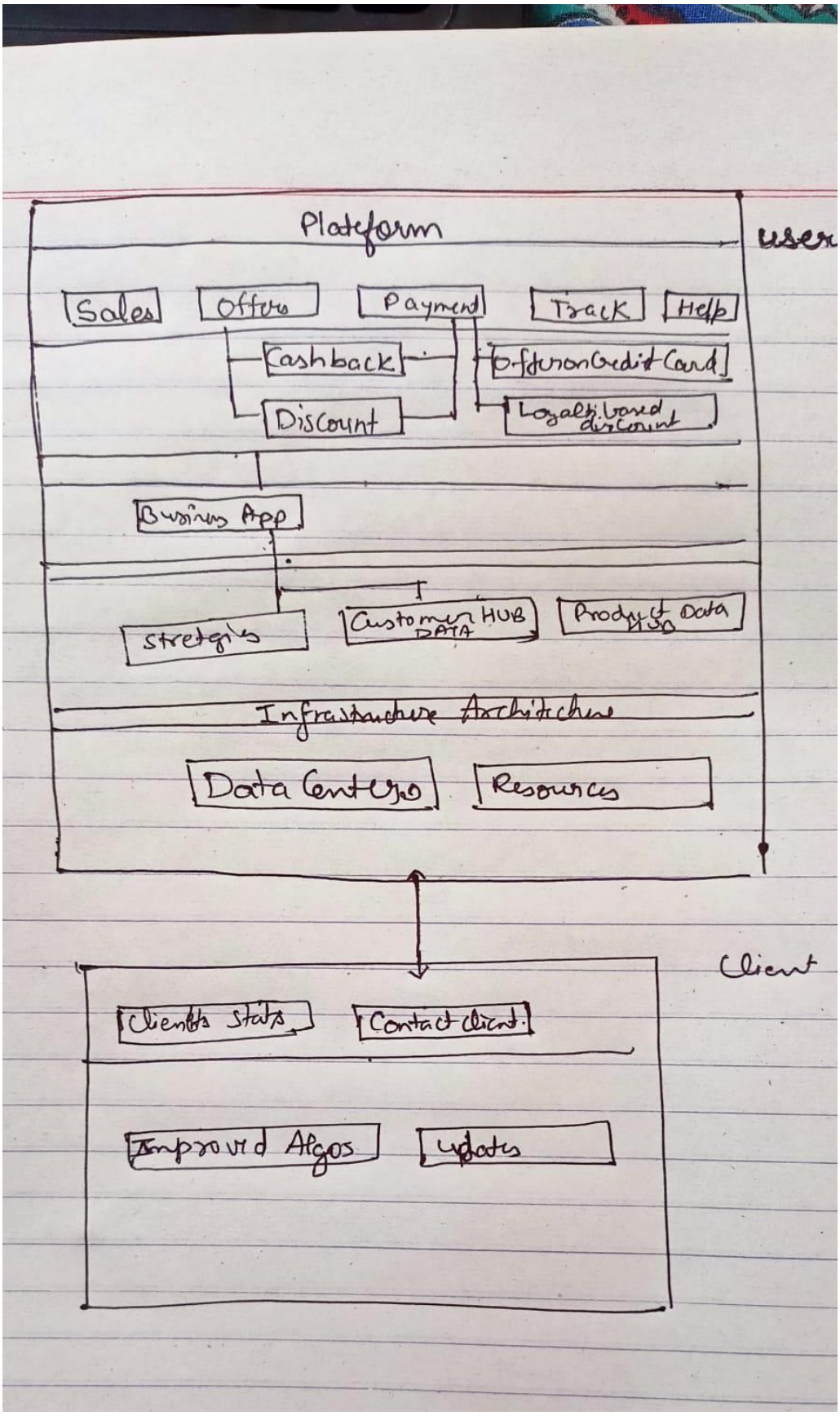
Set(2)

As a product for development Any e-marketing company we need to know the requirement for the company first. Building any applications requires extensive integration with existing info like as corporate databases, intranets, enterprise resource planning (ERP), and other application programs. After learning more about our client's requirement and justify whether it is feasible we can start planning the dummy for our client. As any E-commerce company will require website and a user friendly application. So, gathering a team with good skills in MEAN and MERN and application developer with knowledgeof java or kotlin and by going with plan as SDLC model we can deliver a good project to our client.

1.we can use tableau to first get insight of our product and then after looking at by selecting which product we are going to gain profit. We can add these to our option for selection. Then we can use oracle for resource planning and managing the product offer and set time for the offers too.

2.For layer of system that combine all offer for generic products with which fall under more than 1 offer categories, we can use python with pandas and matplot library. Here we can use pandas to make a aggregate relation with if else conditions and substracting the discounted value and after that adding the cashback of the amount paid by user to the user's wallet .

3.we can use radio buttons for selecting offer if there is multiple options available and can only select one of them.


Plateform
user
Sales
Offers
Payment
Track
Help
Cashback
Offer on Credit Card
Discount
Loyalty based discount
Busines App
Strategies
Customer DATA HUB
Product Data HUB
Infrastructure Architecture
Data Centers
Resources
Client
Clients stats
Contact client
Improved Algos
Updates