

COURSE: EDA, SQL & BIG DATA

SQL MINI PROJECT: E-commerce analysis

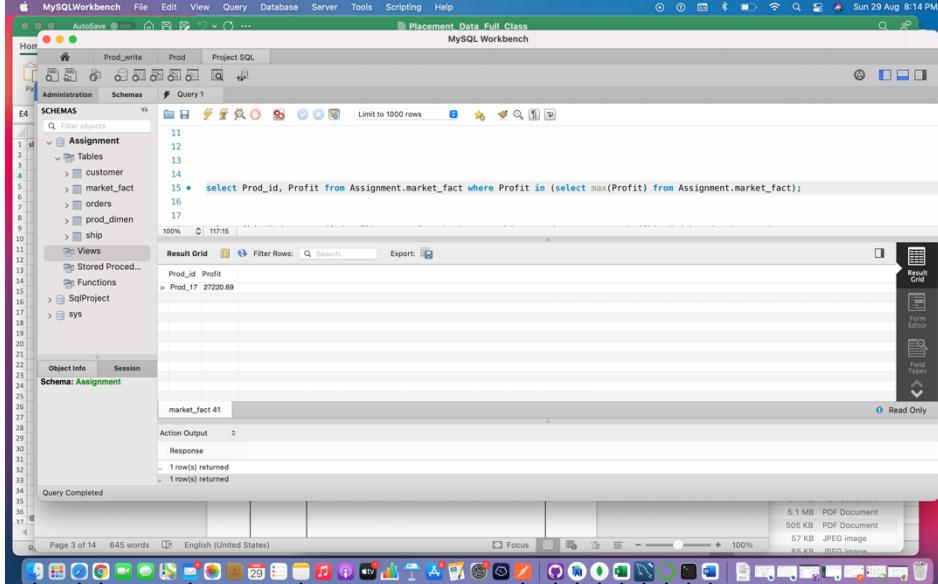
EDA MINI PROJECT: Placement Analysis

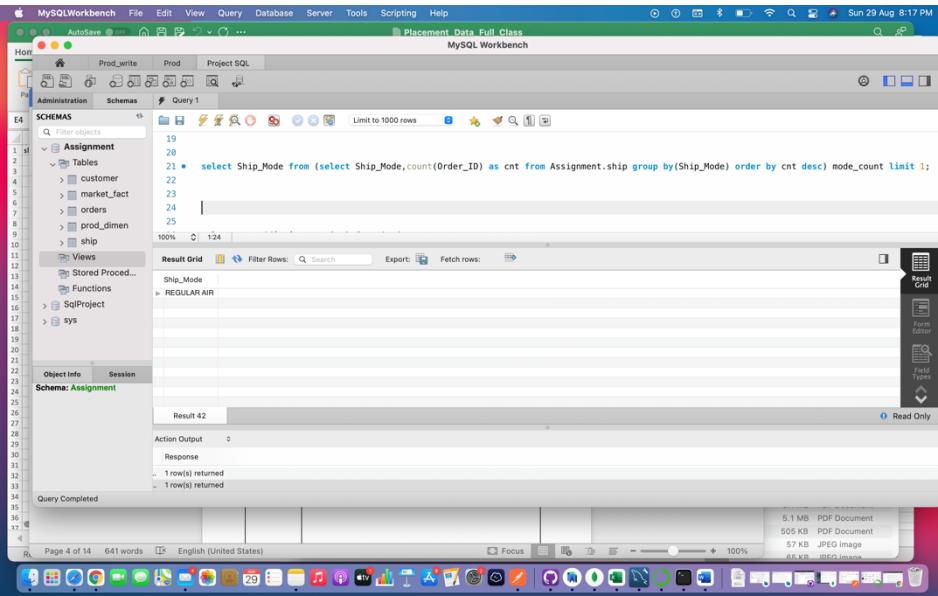
Instructions for the submission:

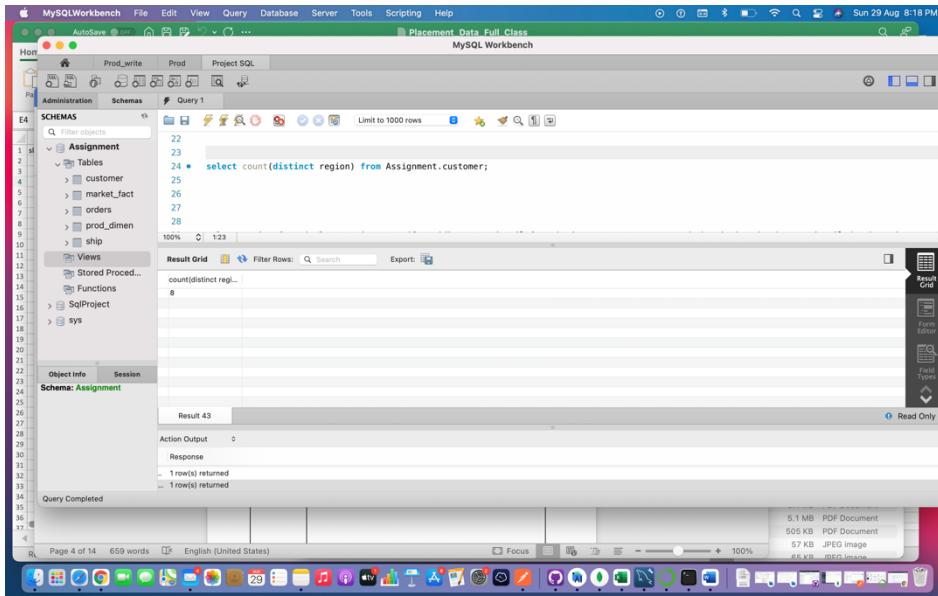
- Please maintain the following: Font - Times New Roman, Font Size - 12, Line Spacing - 1.5
- Wherever mentioned, please ensure that the relevant screenshots of your working from the SQL workbench/Python notebook/Excel Sheet

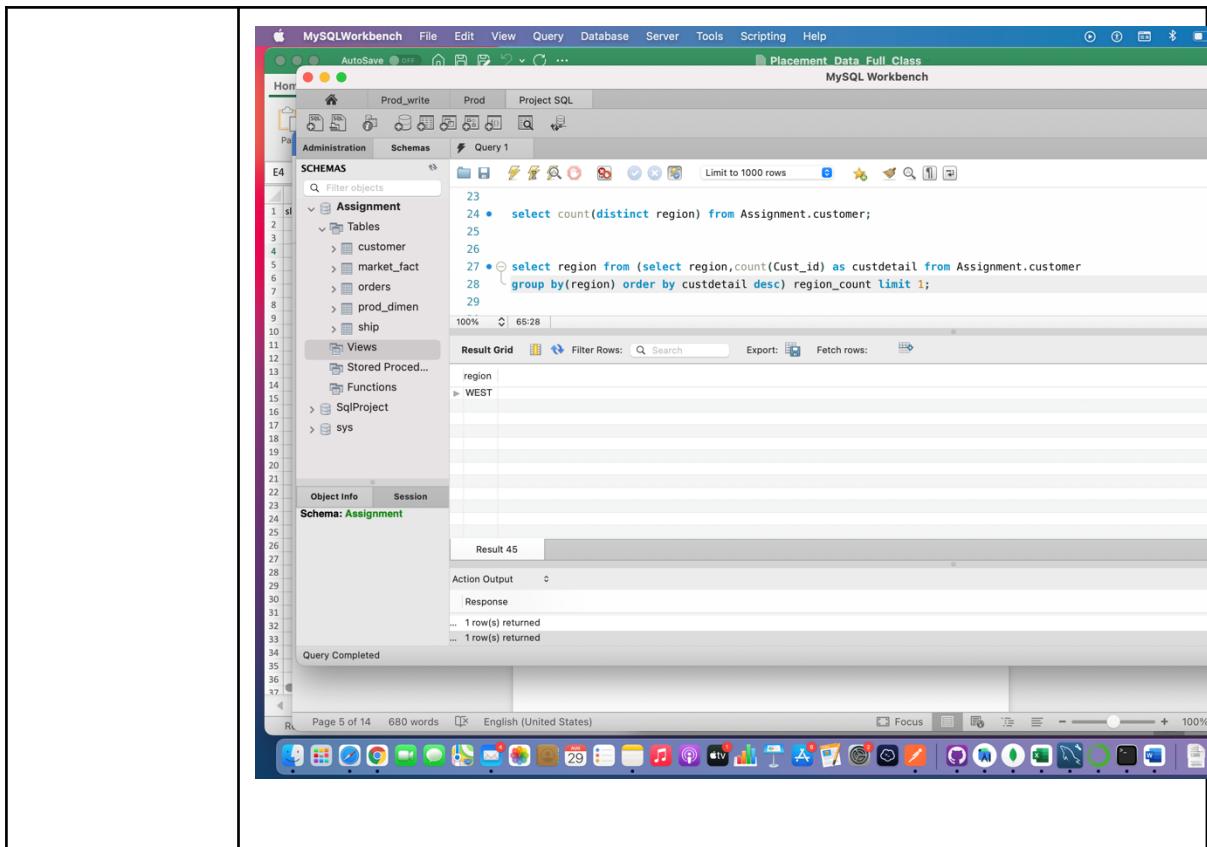
PART I: SQL

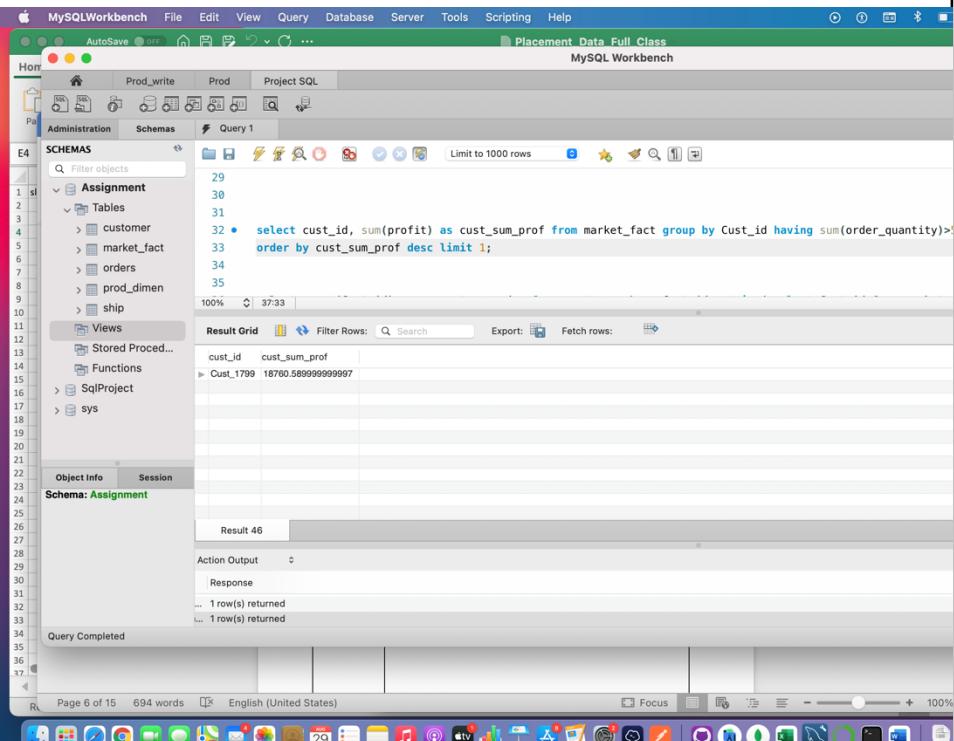
S.No.	Assessment Area	Submission Format	Marks
1	SQL	SQL Query + Screenshot of output	2 marks
2		SQL Query + Screenshot of output	2 marks
3		SQL Query + Screenshot of output	2 marks
4		SQL Query + Screenshot of output	2 marks
5		SQL Query + Screenshot of output	2 marks
6		SQL Query + Screenshot of output	2 marks
7		SQL Query + Screenshot of output	4 marks
8		SQL Query + Screenshot of output	4 marks
Project Part I Maximum Marks			20 marks

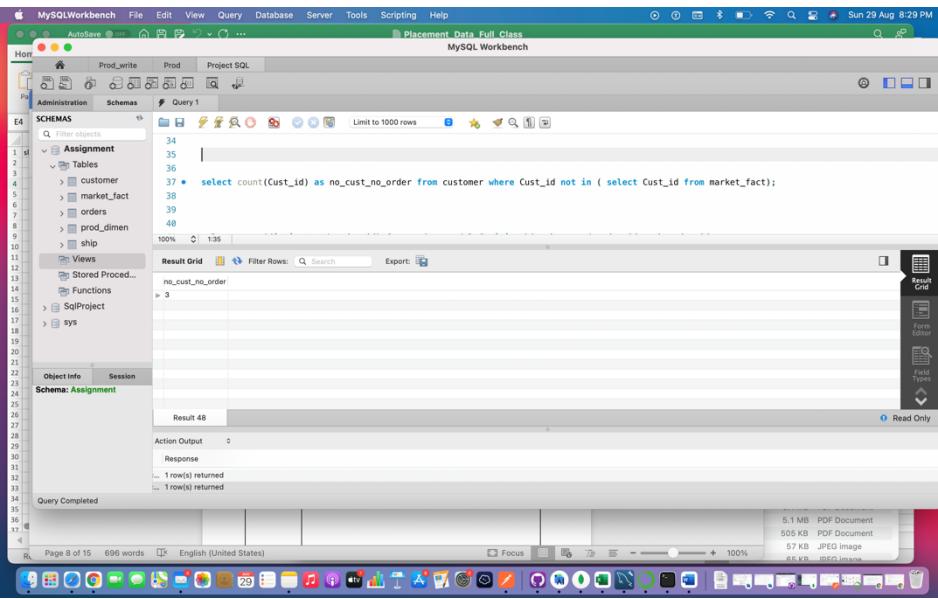
Question 2	You want to focus your marketing and sales resources...				
Marks	2 marks				
SQL Query	<pre>select Prod_id, Profit from Assignment.market_fact where Profit in (select max(Profit) from Assignment.market_fact);</pre>				
Screenshot of query and output	 <p>The screenshot shows the MySQL Workbench interface. The left pane displays the database schema with the 'Assignment' schema selected, containing tables like customer, market_fact, orders, prod_dimen, and ship. The main pane shows a query editor with the following SQL code:</p> <pre>11 12 13 14 15 • select Prod_id, Profit from Assignment.market_fact where Profit in 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36</pre> <p>The result grid shows one row of data:</p> <table border="1"> <thead> <tr> <th>Prod_id</th> <th>Profit</th> </tr> </thead> <tbody> <tr> <td>Prod_17</td> <td>27220.69</td> </tr> </tbody> </table> <p>Below the result grid, the message "Query Completed" is displayed.</p>	Prod_id	Profit	Prod_17	27220.69
Prod_id	Profit				
Prod_17	27220.69				

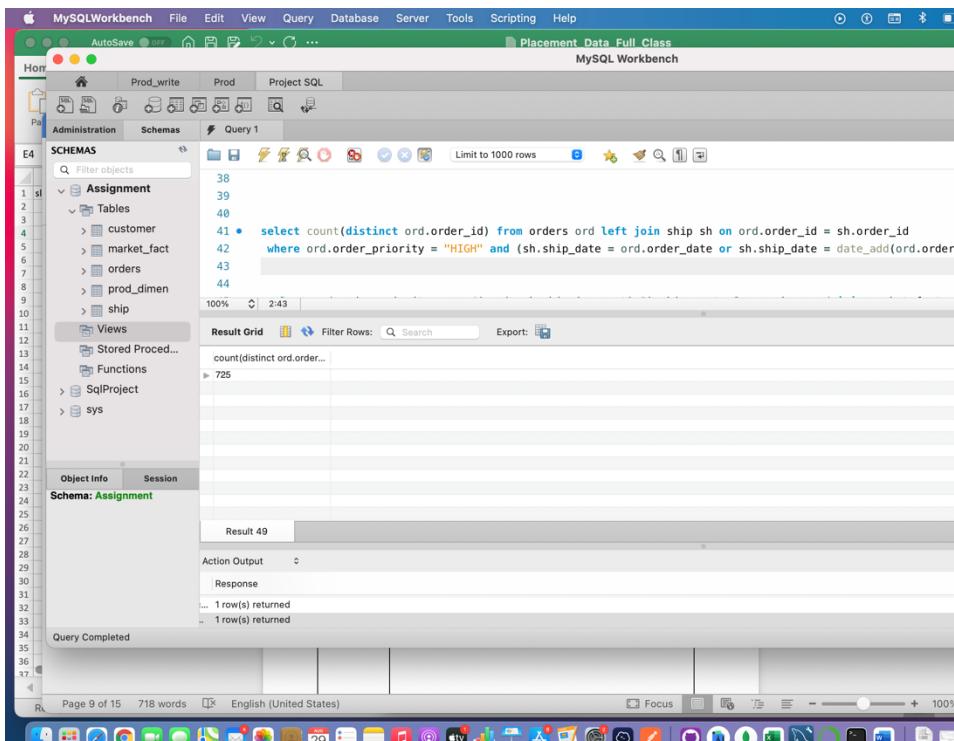
Question 3	Which is the most frequent mode of shipment...		
Marks	2 marks		
SQL Query	<pre>select Ship_Mode from (select Ship_Mode,count(Order_ID) as cnt from Assignment.ship group by(Ship_Mode) order by cnt desc) mode_count limit 1;</pre>		
Screenshot of query and output	<p>Paste your screenshots here</p>  <p>The screenshot shows the MySQL Workbench interface. The left sidebar displays the schema structure for the 'Assignment' database, including tables like customer, market_fact, orders, prod_dimen, and ship. The main window shows a query editor with the following SQL code:</p> <pre>19 20 21 • select Ship_Mode from (select Ship_Mode,count(Order_ID) as cnt from Assignment.ship group by(Ship_Mode) order by cnt desc) mode_count limit 1; 22 23 24 25</pre> <p>The result grid below the query editor shows the output of the query:</p> <table border="1"> <thead> <tr> <th>Ship_Mode</th> </tr> </thead> <tbody> <tr> <td>REGULAR AIR</td> </tr> </tbody> </table> <p>The status bar at the bottom indicates "Query Completed".</p>	Ship_Mode	REGULAR AIR
Ship_Mode			
REGULAR AIR			

Question 4	What is the number of unique regions that the company...
Marks	2 marks
SQL Query	<p>a) select count(distinct region) from Assignment.customer;</p> <p>b) select region from (select region,count(Cust_id) as custdetail from Assignment.customer group by(region) order by custdetail desc) region_count limit 1;</p>
Screenshot of query and output	<p>Paste your screenshots here</p>  <pre> 22 23 24 • select count(distinct region) from Assignment.customer; 25 26 27 28 Result Grid count(distinct regi... 8 Action Output Response - 1 row(s) returned - 1 row(s) returned Query Completed </pre>



Question 5	To improve your customer retention and give them...				
Marks	2 marks				
SQL Query	<pre>select cust_id, sum(profit) as cust_sum_prof from market_fact group by Cust_id having sum(order_quantity)>500 order by cust_sum_prof desc limit 1;</pre>				
Screenshot of query and output	 <p>The screenshot shows the MySQL Workbench interface. The left pane displays the database schema with tables like customer, market_fact, orders, prod_dimen, and ship. The central pane shows the SQL editor with the following query:</p> <pre> 29 30 31 32 • select cust_id, sum(profit) as cust_sum_prof from market_fact group by Cust_id having sum(order_quantity)> 33 order by cust_sum_prof desc limit 1; 34 35 </pre> <p>The result grid shows one row of data:</p> <table border="1"> <thead> <tr> <th>cust_id</th> <th>cust_sum_prof</th> </tr> </thead> <tbody> <tr> <td>Cust_1799</td> <td>18780.589999999997</td> </tr> </tbody> </table> <p>Below the result grid, the message "1 row(s) returned" is displayed.</p>	cust_id	cust_sum_prof	Cust_1799	18780.589999999997
cust_id	cust_sum_prof				
Cust_1799	18780.589999999997				

Question 6	Find the number of customers who have not made any...		
Marks	2 marks		
SQL Query	<pre>select count(Cust_id) as no_cust_no_order from customer where Cust_id not in (select Cust_id from market_fact);</pre>		
Screenshot of query and output	 <p>The screenshot shows the MySQL Workbench interface. The left pane displays the database schema with tables like customer, market_fact, orders, prod_dimen, and ship. The right pane shows the SQL editor with the following query:</p> <pre>34 35 36 37 • select count(Cust_id) as no_cust_no_order from customer where Cust_id not in (select Cust_id from market_fact); 38 39 40</pre> <p>The result grid shows the output of the query:</p> <table border="1"> <thead> <tr> <th>no_cust_no_order</th> </tr> </thead> <tbody> <tr> <td>3</td> </tr> </tbody> </table> <p>Below the result grid, the status bar indicates "Query Completed".</p>	no_cust_no_order	3
no_cust_no_order			
3			

Question 7	According to the e-commerce company, a High priority order...
Marks	4 marks
SQL Query	<p>a) select count(distinct ord.order_id) from orders ord left join ship sh on ord.order_id = sh.order_id where ord.order_priority = "HIGH" and (sh.ship_date = ord.order_date or sh.ship_date = date_add(ord.order_date, interval 1 day));</p> <p>b) select count(distinct ord.order_id) from orders ord left join ship sh on ord.order_id = sh.order_id where ord.order_priority = "HIGH";</p>
Screenshot of query and output	 <pre> MySQLWorkbench File Edit View Query Database Server Tools Scripting Help AutoSave OFF Home Project SQL Placement Data Full Class MySQL Workbench Schemas E4 1 sl 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 SCHEMAS Filter objects Assignment Tables customer market_fact orders prod_dimen ship Views Stored Proced... Functions SqlProject sys Object Info Session Schema: Assignment Result Grid Filter Rows: Search Export: count(distinct ord.order... 725 Action Output Response ... 1 row(s) returned ... 1 row(s) returned Query Completed </pre>

```

35
36
37
38
39 • select count(distinct ord.order_id) from orders ord left join ship sh on ord.order_id = sh.order_id
40   where ord.order_priority = "HIGH" and (sh.ship_date = ord.order_date or sh.ship_date = date_add(ord.order_date, interval 1 day));
41
42 • select count(distinct ord.order_id) from orders ord left join ship sh on ord.order_id = sh.order_id
43   where ord.order_priority = "HIGH";
44
45
46
47

```

Result Grid

count(distinct ord.order_id)
1137

Action Output

- Response
- 1row(s) returned
- 1row(s) returned

Query Completed

The claim made by the company is false because total high priority orders are 1137 but only 735 high priority orders are delivered on the same date or next date of ordered date.

Question 8	Compare the average shipping cost for various “order_priority” ...
Marks	4 marks
SQL Query	<pre> select ord.order_priority, round(avg(mark.shipping_cost),2) ship_cost from orders ord join market_fact mark on ord.order_id = mark.order_id join customer cust on cust.cust_id = mark.cust_id where cust.region = "west" group by order_priority ; </pre>

Screenshot of query and output

```

43
44
45
46 •  select ord.order_priority, round(avg(mark.shipping_cost),2) ship_cost  from orders ord join market_fact mark on ord.ord_id = mark.ord_id
47   join customer cust on cust.cust_id = mark.cust_id where cust.region = "west" group by order_priority ;
48
49

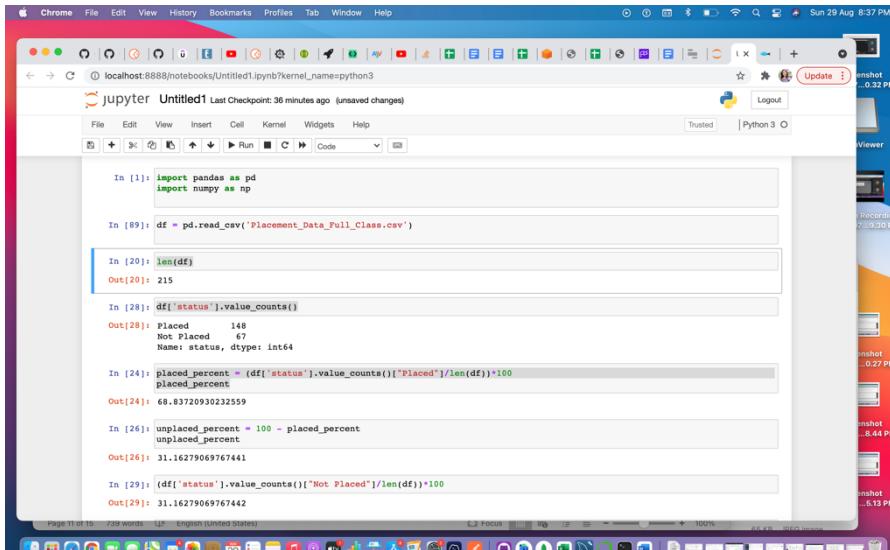
```

order_priority	ship_cost
CRITICAL	11.98
LOW	14.13
NOT SPECIFIED	10.79
HIGH	11.57
MEDIUM	12.67

PART II: EDA

S.No .	Assessment Area	Submission Format	Marks
1	EDA	Steps taken/Python code + Screenshot of output	3marks
2		Steps taken/Python code + Screenshot of output	3 marks
3		Steps taken/Python code + Screenshot of output	3 marks

4		Steps taken/Python code + Screenshot of output	3 marks
5		Steps taken/Python code + Screenshot of output	8 marks
Project Part II Maximum Marks		20 marks	

Question 1	What is the percentage of people placed
Marks	3 Marks
Solution	<p>Steps taken :-</p> <ol style="list-style-type: none"> 1) Calculate the length of dataframe i.e. total no. of students. 2) Used .value_counts() to find out no. of “Placed” and “Not placed” students. 3) Calculated the percentage for the above two by status type diving by total no. of students and multiplying by 100.  <pre> In [1]: import pandas as pd import numpy as np In [89]: df = pd.read_csv('Placement_Data_Full_Class.csv') In [20]: len(df) Out[20]: 215 In [28]: df['status'].value_counts() Out[28]: Placed 148 Not Placed 67 Name: status, dtype: int64 In [24]: placed_percent = (df['status'].value_counts()['Placed']/len(df))*100 placed_percent Out[24]: 68.83720930232559 In [26]: unplaced_percent = 100 - placed_percent unplaced_percent Out[26]: 31.16279069767441 In [29]: (df['status'].value_counts()['Not Placed']/len(df))*100 Out[29]: 31.16279069767442 </pre>

Question 2	Find the minimum, maximum, median and average.....
Marks	3 Marks

Solution

Steps taken:-

- 1) Use df['salary'].min() to calculate minimum salary.
- 2) Use df['salary'].max() to calculate maximum salary.
- 3) Similarly used median and mean function to calculate median and average salary offered to students

Yes, the average is a good measure of central tendency of data.

The screenshot shows a Jupyter Notebook running in a browser window. The code cells and their outputs are as follows:

```
In [30]: df['salary'].min()
Out[30]: 200000.0

In [31]: df['salary'].max()
Out[31]: 940000.0

In [33]: df['salary'].median(skipna = True)
Out[33]: 265000.0

In [36]: df['salary'].mean(skipna = True)
Out[36]: 288655.4054054054

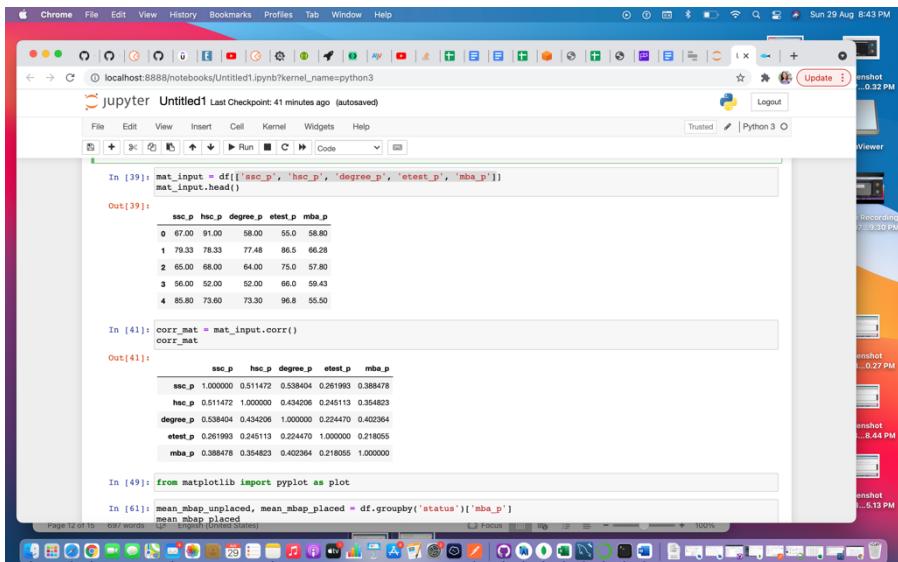
In [39]: mat_input = df[['ssc_p', 'hsc_p', 'degree_p', 'etest_p', 'mba_p']]
mat_input.head()

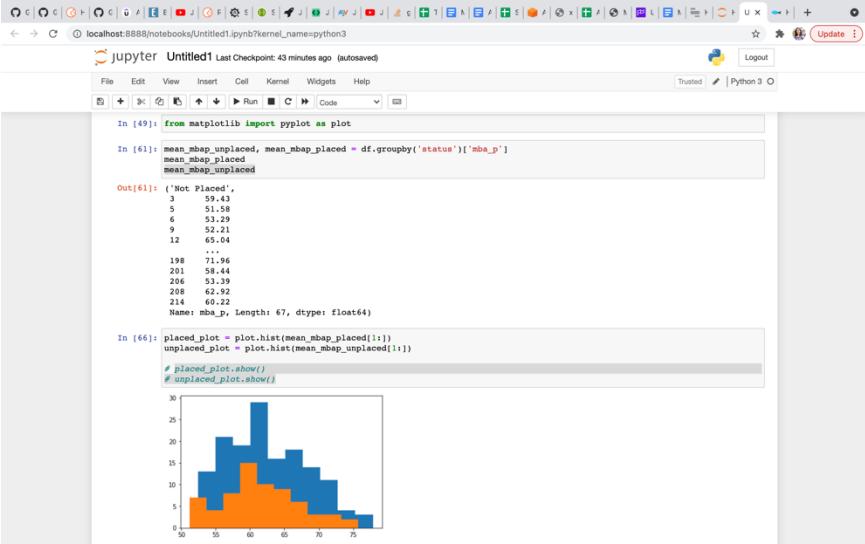
Out[39]:
   ssc_p  hsc_p  degree_p  etest_p  mba_p
0    67.00    91.00     58.00    55.00    58.80
1    79.33    78.33     77.48    86.5    66.28
2    65.00    68.00     64.00    75.0    57.80
3    56.00    52.00     52.00    66.0    59.43
4    85.80    73.60     73.30    96.8    55.50
```

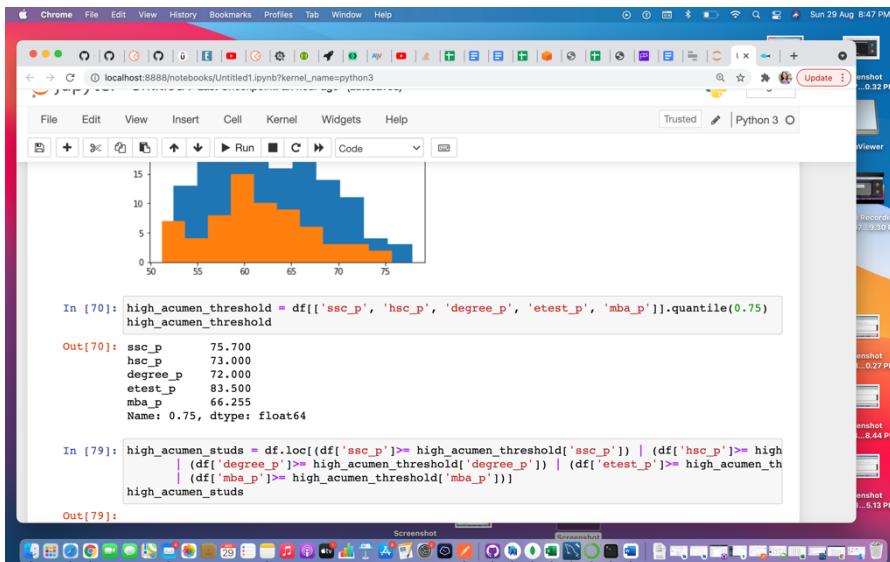
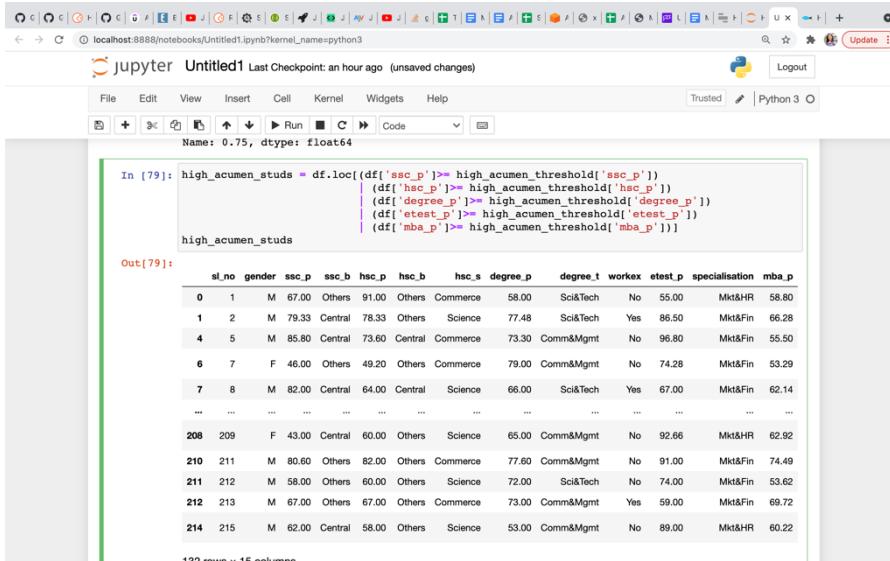
In [41]: corr_mat = mat_input.corr()

Question 3

Analyse the correlation between the following

Marks	3 Marks
Solution	<p>Steps taken :-</p> <ol style="list-style-type: none"> 1) Extracted the required columns from the dataframe. 2) Used .corr() to calculate the correlation matrix between all the pairs of columns. <p>My comment: The correlation between these tests is not very prominent as they lie between 20% to 55% only.</p> <p>Highest Correlation: degree_p and ssc_p Lowest Correlation: mba_p and etest_p</p>  <pre> In [39]: mat_input = df[['ssc_p', 'hsc_p', 'degree_p', 'etest_p', 'mba_p']] mat_input.head() Out[39]: ssc_p hsc_p degree_p etest_p mba_p 0 67.00 91.00 58.00 55.0 58.80 1 79.33 78.33 77.48 86.5 62.28 2 65.00 68.00 64.00 75.0 57.80 3 56.00 52.00 52.00 68.0 59.43 4 85.80 73.60 73.30 96.8 55.50 In [41]: corr_mat = mat_input.corr() corr_mat Out[41]: ssc_p hsc_p degree_p etest_p mba_p ssc_p 1.00000 0.511472 0.538404 0.261993 0.388478 hsc_p 0.511472 1.000000 0.434206 0.245113 0.354823 degree_p 0.538404 0.434206 1.000000 0.224470 0.402364 etest_p 0.261993 0.245113 0.224470 1.000000 0.218055 mba_p 0.388478 0.354823 0.402364 0.218055 1.000000 In [49]: from matplotlib import pyplot as plot In [51]: mean_mba_p_uploaded, mean_mba_p_placed = df.groupby('status')[['mba_p']] mean_mba_p_uploaded mean_mba_p_placed </pre>

Question 4	You want to compare the MBA_P score.....
Marks	3 Marks
Solution	<p>On analysis of the plots we can say that there is no such impact of mba_p score and placement of students. This is because the pattern for the number for students scoring a certain marks in mba_p test is same for “Placed” as well as “Not Placed” students, with majority lying in the range of 57 to 63 marks.</p>  <pre>In [49]: from matplotlib import pyplot as plot In [61]: mean_mbap_unplaced, mean_mbap_placed = df.groupby('status')[['mba_p']] mean_mbap_placed mean_mbap_unplaced Out[61]: ('Not Placed', 3 59.43 5 59.48 6 53.29 9 52.21 12 65.04 .. 198 71.96 201 58.44 206 53.39 208 62.52 214 60.22 Name: mba_p, Length: 67, dtype: float64) In [66]: placed_plot = plot.hist(mean_mbap_placed[1:]) unplaced_plot = plot.hist(mean_mbap_unplaced[1:]) # placed_plot.show() # unplaced_plot.show()</pre> <p>The screenshot shows a Jupyter Notebook interface with a histogram titled 'placed_plot'. The x-axis represents the 'mba_p' score, ranging from 50 to 75. The y-axis represents frequency, ranging from 0 to 30. The histogram consists of two overlapping distributions: one in blue (representing 'Placed' students) and one in orange (representing 'Unplaced' students). Both distributions are centered around a score of approximately 60, with a slight peak between 55 and 65. The total length of the data is 67.</p>

Question 5	You want to understand the impact of academic																																																																																																																																																																	
Marks	8 Marks																																																																																																																																																																	
5.1	 <p>A screenshot of a Jupyter Notebook interface. At the top, there's a toolbar with various icons. Below it is a menu bar with 'File', 'Edit', 'View', etc. A status bar at the bottom shows 'Sun 29 Aug 8:47 PM'. The main area contains a histogram with blue and orange bars, followed by several code cells. The first cell shows:</p> <pre>In [70]: high_acumen_threshold = df[['ssc_p', 'hsc_p', 'degree_p', 'etest_p', 'mba_p']].quantile(0.75) high_acumen_threshold</pre> <p>The output shows the quantiles for each column:</p> <pre>Out[70]: ssc_p 75.700 hsc_p 73.000 degree_p 72.000 etest_p 83.500 mba_p 66.255 Name: 0.75, dtype: float64</pre> <p>The second cell shows:</p> <pre>In [79]: high_acumen_studs = df.loc[(df['ssc_p']>= high_acumen_threshold['ssc_p']) (df['hsc_p']>= high_acumen_threshold['hsc_p']) (df['degree_p']>= high_acumen_threshold['degree_p']) (df['etest_p']>= high_acumen_threshold['etest_p']) (df['mba_p']>= high_acumen_threshold['mba_p'])] high_acumen_studs</pre> <p>The output shows the resulting DataFrame:</p> <pre>Out[79]:</pre>																																																																																																																																																																	
5.2	 <p>A screenshot of a Jupyter Notebook interface. The title bar says 'jupyter Untitled1 Last Checkpoint: an hour ago (unsaved changes)'. The main area shows a large DataFrame output. The first few rows of the DataFrame are:</p> <table border="1"> <thead> <tr> <th>sl_no</th> <th>gender</th> <th>ssc_p</th> <th>ssc_b</th> <th>hsc_p</th> <th>hsc_b</th> <th>hsc_s</th> <th>degree_p</th> <th>degree_t</th> <th>workex</th> <th>etest_p</th> <th>specialisation</th> <th>mba_p</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>M</td> <td>67.00</td> <td>Others</td> <td>91.00</td> <td>Others</td> <td>Commerce</td> <td>58.00</td> <td>Sci&Tech</td> <td>No</td> <td>55.00</td> <td>Mkt&HR</td> <td>58.80</td> </tr> <tr> <td>1</td> <td>M</td> <td>79.33</td> <td>Central</td> <td>78.33</td> <td>Others</td> <td>Science</td> <td>77.48</td> <td>Sci&Tech</td> <td>Yes</td> <td>86.50</td> <td>Mkt&Fin</td> <td>66.28</td> </tr> <tr> <td>4</td> <td>M</td> <td>85.80</td> <td>Central</td> <td>73.60</td> <td>Central</td> <td>Commerce</td> <td>73.30</td> <td>Comm&Mgmt</td> <td>No</td> <td>96.80</td> <td>Mkt&Fin</td> <td>55.50</td> </tr> <tr> <td>6</td> <td>F</td> <td>46.00</td> <td>Others</td> <td>49.20</td> <td>Others</td> <td>Commerce</td> <td>79.00</td> <td>Comm&Mgmt</td> <td>No</td> <td>74.28</td> <td>Mkt&Fin</td> <td>53.29</td> </tr> <tr> <td>7</td> <td>M</td> <td>82.00</td> <td>Central</td> <td>64.00</td> <td>Central</td> <td>Science</td> <td>66.00</td> <td>Sci&Tech</td> <td>Yes</td> <td>67.00</td> <td>Mkt&Fin</td> <td>62.14</td> </tr> <tr> <td>...</td> </tr> <tr> <td>208</td> <td>209</td> <td>F</td> <td>43.00</td> <td>Central</td> <td>60.00</td> <td>Others</td> <td>Science</td> <td>65.00</td> <td>Comm&Mgmt</td> <td>No</td> <td>92.66</td> <td>Mkt&HR</td> <td>62.92</td> </tr> <tr> <td>210</td> <td>211</td> <td>M</td> <td>80.60</td> <td>Others</td> <td>82.00</td> <td>Others</td> <td>Commerce</td> <td>77.60</td> <td>Comm&Mgmt</td> <td>No</td> <td>91.00</td> <td>Mkt&Fin</td> <td>74.49</td> </tr> <tr> <td>211</td> <td>212</td> <td>M</td> <td>58.00</td> <td>Others</td> <td>60.00</td> <td>Others</td> <td>Science</td> <td>72.00</td> <td>Sci&Tech</td> <td>No</td> <td>74.00</td> <td>Mkt&Fin</td> <td>53.62</td> </tr> <tr> <td>212</td> <td>213</td> <td>M</td> <td>67.00</td> <td>Others</td> <td>67.00</td> <td>Others</td> <td>Commerce</td> <td>73.00</td> <td>Comm&Mgmt</td> <td>Yes</td> <td>59.00</td> <td>Mkt&Fin</td> <td>69.72</td> </tr> <tr> <td>214</td> <td>215</td> <td>M</td> <td>62.00</td> <td>Central</td> <td>58.00</td> <td>Others</td> <td>Science</td> <td>53.00</td> <td>Comm&Mgmt</td> <td>No</td> <td>89.00</td> <td>Mkt&HR</td> <td>60.22</td> </tr> </tbody> </table> <p>Below the table, it says '192 rows x 15 columns'.</p>	sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p	0	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.00	Mkt&HR	58.80	1	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.50	Mkt&Fin	66.28	4	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.80	Mkt&Fin	55.50	6	F	46.00	Others	49.20	Others	Commerce	79.00	Comm&Mgmt	No	74.28	Mkt&Fin	53.29	7	M	82.00	Central	64.00	Central	Science	66.00	Sci&Tech	Yes	67.00	Mkt&Fin	62.14	208	209	F	43.00	Central	60.00	Others	Science	65.00	Comm&Mgmt	No	92.66	Mkt&HR	62.92	210	211	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.00	Mkt&Fin	74.49	211	212	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.00	Mkt&Fin	53.62	212	213	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.00	Mkt&Fin	69.72	214	215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.00	Mkt&HR	60.22
sl_no	gender	ssc_p	ssc_b	hsc_p	hsc_b	hsc_s	degree_p	degree_t	workex	etest_p	specialisation	mba_p																																																																																																																																																						
0	M	67.00	Others	91.00	Others	Commerce	58.00	Sci&Tech	No	55.00	Mkt&HR	58.80																																																																																																																																																						
1	M	79.33	Central	78.33	Others	Science	77.48	Sci&Tech	Yes	86.50	Mkt&Fin	66.28																																																																																																																																																						
4	M	85.80	Central	73.60	Central	Commerce	73.30	Comm&Mgmt	No	96.80	Mkt&Fin	55.50																																																																																																																																																						
6	F	46.00	Others	49.20	Others	Commerce	79.00	Comm&Mgmt	No	74.28	Mkt&Fin	53.29																																																																																																																																																						
7	M	82.00	Central	64.00	Central	Science	66.00	Sci&Tech	Yes	67.00	Mkt&Fin	62.14																																																																																																																																																						
...																																																																																																																																																						
208	209	F	43.00	Central	60.00	Others	Science	65.00	Comm&Mgmt	No	92.66	Mkt&HR	62.92																																																																																																																																																					
210	211	M	80.60	Others	82.00	Others	Commerce	77.60	Comm&Mgmt	No	91.00	Mkt&Fin	74.49																																																																																																																																																					
211	212	M	58.00	Others	60.00	Others	Science	72.00	Sci&Tech	No	74.00	Mkt&Fin	53.62																																																																																																																																																					
212	213	M	67.00	Others	67.00	Others	Commerce	73.00	Comm&Mgmt	Yes	59.00	Mkt&Fin	69.72																																																																																																																																																					
214	215	M	62.00	Central	58.00	Others	Science	53.00	Comm&Mgmt	No	89.00	Mkt&HR	60.22																																																																																																																																																					

5.3

The screenshot shows a Jupyter Notebook interface running in a Chrome browser. The notebook has a single open cell labeled 'In [86]'. The code in the cell is:

```
In [86]: high_acumen_studs['status'].value_counts()
```

The output of the cell, labeled 'Out[86]', is:

```
Out[86]: Placed      106  
Not Placed    26  
Name: status, dtype: int64
```

The next cell, 'In [87]', contains the following code:

```
In [87]: HA_placed_percent = (high_acumen_studs['status'].value_counts()['Placed']/len(high_acumen_studs))  
HA_placed_percent
```

The output of 'In [87]', labeled 'Out[87]', is:

```
Out[87]: 80.30303030303030
```

The final cell, 'In [88]', contains:

```
In [88]: HA_unplaced_percent = (high_acumen_studs['status'].value_counts()['Not Placed']/len(high_acumen_studs))  
HA_unplaced_percent
```

The output of 'In [88]', labeled 'Out[88]', is:

```
Out[88]: 19.696969696969695
```

The browser window has a tab bar at the top with various icons. On the right side of the browser window, there is a sidebar showing a list of previous screenshots taken at different times.

5.4

Steps Taken: -

- 1) We calculated the 75th percentile value for all the required values by using .quantile(0.75)
- 2) Then we filtered all whole dataframe based on the 75th percentile values of the test scores. If a student has score marks equal to or above the 75th percentile marks for that test, he/she is considered as student of High Acumen.
- 3) We calculated the placement percentage for the students with high acumen.

Since, among the high acumen students, 80 % students got placed we can say that the hypothesis is somewhat true i.e. the high acumen students mostly leads to better placements.