# FINAL EXAM PAPER

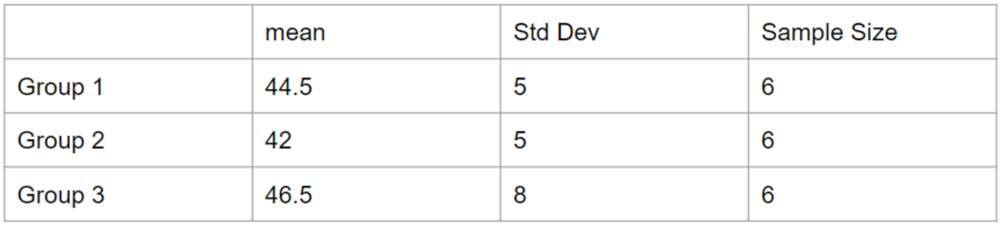
**[Time: 2 hrs.]**

**[Total Marks: 100]**

**Instructions:**

- Attempt all the questions
- **All the answers will be checked for plagiarism. Any incident of copied content from the internet will result in negative marking.**
- Marks of the questions are written along with the questions
- Pls submit your answers with the naming convention “PGA-INT_COLLEGE_Batch#_<Student Name>”

| Sr.No. | Questions | Marks |
|---|---|---|
| **Q1.** | Print customer_id, account_number, balance_amount, conPrint account_number, balance_amount, transaction_amount from Bank_Account_Details and bank_account_transaction for all the transactions occurred during march, 2020 and april, 2020<br><br>Refer SQL Dataset | **[5]** |
| **Q2.** | Write a program to fetch the words from the given list which have their first character in uppercase.<br>namesList = ['santa Maria', 'Hello World','Merry christmas', 'tHank You']<br>Output: ['Maria', 'Hello', 'World', 'Merry', 'You'] | **[10]** |
| **Q3.** | A list of dates (dd-mm-yyyy) in the form of string is given below. Create a new list that stores years i.e. the 'yyyy' part from the dates in the given list.<br><br>datesList = ['17-12-1997','22-04-2011','01-05-1993','19-06-2020']<br>Output: ['1997','2011','1993','2020'] | **[10]** |
| **Q4.** | Annual project of university done by three groups of students with equal sample sizes. Each group was given a different task. After the final presentation students get the marks. The summary statistics are given below: | **[10]** |



| | mean | Std Dev | Sample Size |
|---|---|---|---|
| Group 1 | 44.5 | 5 | 6 |
| Group 2 | 42 | 5 | 6 |
| Group 3 | 46.5 | 8 | 6 |

Calculate F - statistics

**Q5.** Find the fruits that are sour in taste from the tuple given below.

**[10]**

```
fruits = (('Lemon','sour'),
          ('DragonFruit', 'Sweet'),
          ('Grapes','soUr'),
          ('Kiwi','Sour'),
          ('Apples','sweet'),
          ('Orange','sour'),
          ('Blueberries','sweet'),
          ('Limes','Sour')
 )
Output:
Sour Fruits:  ['Lemon', 'Grapes', 'Kiwi', 'Orange', 'Limes']
```

**Q6.** Write a query to determine the Nth (Say N=5) highest purch_amt from Orders table.

**[10]**

Refer SQL Dataset

**Q7.** The data given contains the information about Uber's ridership.

**[10]**

Dataset Information:
Request ID: Ride request ID
Date: Date of request
Request time: Time at which the ride was requested
Dropoff time: Time at which the ride was completed
Pickup Point: Customer's pickup spot
Driver ID: Driver's unique identification number
Status: Ride status

Perform the following tasks using Tableau with the aim to identify patterns which indicate:

1. The major time slots when the demand is more

Refer Tableau Dataset

**Q8.** The number of deaths in the 7 metro cities of the US during the last month due to Covid-19 is recorded by the government, and that time, experts predicted that upto next month death rate will be increased by

**[10]**

| | | |
|---|---|---|
| | 30%. So, the number of deaths in current month is also recorded in 7 cities in US are given below:<br>City = ["New York", "New Jersey", "Michigan", "California", "Florida", "Massachusetts", "Texas"]<br>No_of_deaths_in_last_month = [3406, 1469, 662, 583, 582, 526, 461]<br>No_of_deaths_in_current_month = [4398, 1846, 1288, 382, 879, 430, 321]<br>df = pd.DataFrame({"City": City, "No_of_deaths_in_last_month":No_of_deaths_in_last_month, "No_of_deaths_in_current_month": No_of_deaths_in_current_month})<br>df<br>Is the death distribution of the current month the same as the expert's prediction? Use the level of significance is 0.1. | |
| **Q9.** | Insert five rows into the cast table where the ids for movie should be 936,939,942,930,941 and their respective roles should be Darth Vader, Sarah Connor, Ethan Hunt, Travis Bickle, Antoine Doinel & their actor ids should be set up as 126,140,135,131,144.<br>Refer SQL Dataset | **[5]** |
| **Q10.** | Given the 'credit_card' dataset, below is the data definition:<br>1) CUSTID: Identification of Credit Card holder (Categorical)<br>2) BALANCE: Balance amount left in their account to make purchases<br>3) BALANCEFREQUENCY: How frequently the Balance is updated, score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)<br>4) PURCHASES: Amount of purchases made from account<br>5) ONEOFFPURCHASES: Maximum purchase amount done in one-go<br>6) INSTALLMENTSPURCHASES: Amount of purchase done in installment<br>7) CASHADVANCE: Cash in advance given by the user<br>8) PURCHASESFREQUENCY: How frequently the Purchases are being made, score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased) | **[10]** |

| | | |
|---|---|---|
| | 9) ONEOFFPURCHASESFREQUENCY: How frequently Purchases are happening in one-go (1 = frequently purchased, 0 = not frequently purchased)<br>10) PURCHASESINSTALLMENTSFREQUENCY: How frequently purchases in installments are being done (1 = frequently done, 0 = not frequently done)<br>11) CASHADVANCEFREQUENCY: How frequently the cash in advance being paid<br>12) CASHADVANCETRX: Number of Transactions made with "Cash in Advanced"<br>13) PURCHASESTRX: Number of purchase transactions made<br>14) CREDITLIMIT: Limit of Credit Card for user<br>15) PAYMENTS: Amount of Payment done by user<br>16) MINIMUM_PAYMENTS: Minimum amount of payments made by user<br>17) PRCFULLPAYMENT: Percent of full payment paid by user<br>18) TENURE: Tenure of credit card service for user<br>Perform the following Exploratory Data Analysis tasks:<br>a. Missing Value Analysis<br>b. Outlier Treatment using the Z-score method<br>c. Deal with correlated variables | |
| **Q11.** | Given is the boston.csv dataset with the following variable information:<br># CRIM - Per Capita crime rate<br># ZN - Proportion of residential land zoned for lots over 25000 sq. ft # INDUS - Proportion of non-retial business acres<br># CHAS - Charles River dummy variable (1 - if tracts bounds river, 0 - otherwise) # NOX - Nitrogen Oxide concentration<br># RM - Average number of rooms per dwelling<br># AGE - Proportion of owner-occupied unit built prior 1940<br># DIS - Weighted MEan of distances of five Boston Employement Centres<br># RAD - Index of accessibilities to Radial highways<br># TAX - Full-value-property-tax rates per $10,000 # PT - Pupil-teacher Ratio<br># B - the proportion of blacks | **[10]** |

| | | |
|---|---|---|
| | # LSTAT - Lower Status of the Population (%)<br># MV - Median Value of homes (Target Variable)<br><br>Read the data from Hive table as spark dataframe | |