UTHM
Universiti Tun Hussein Onn Malaysia

UNIVERSITI TUN HUSSEIN ONN MALAYSIA

FINAL EXAMINATION
(ONLINE)
SEMESTER II
SESSION 2020/2021

COURSE NAME	:	DATA SCIENCE AND APPLICATIONS
COURSE CODE	:	BFS 41203
PROGRAMME CODE	:	BFF
EXAMINATION DATE	:	JULY 2021
DURATION	:	3 HOURS
INSTRUCTION	:	ANSWER **ALL** QUESTIONS

THIS QUESTION PAPER CONSISTS OF **FIVE (5)** PAGES

Q1 (a) Discuss **SIX (6)** phases of *Data Analytic Lifecycles* with appropriate case-study in the field of civil engineering.

(18 marks)

(b) List and elaborate main analytics spectrum in Data Science using example of traffic engineering.

(7 marks)

Q2 (a) The series of value store are given below:

```
data1 = [10,15,16,18,27]
data2 = [102,118,127,132,150]
data3 = ['a','b','c','d','e']
```

(i) Write a complete Python's code to create a *Pandas DataFrame* and assign it into a variable named as `df`. Import appropriate library.

(6 marks)

(ii) Update your code in **a(i)** to assign `data3` as index.

(3 marks)

(iii) Write a Python's code to show the linear relationship between variable `data1` and `data2`.

(6 marks)

(b) **Figure Q2** shows the snapshot of imported dataset into Jupyter Notebook as dataframe `df`. Write Python's code to perform the following data exploration requirement.

(i) Compute each value of measure of central tendency for variable `x1`.

(ii) Identify missing values of the dataset `df`.

(iii) Calculate the frequency for each class in variable `y`.

(iv) Replace value of 1 with YES and 0 with NO in the variable `x2`.

(v) Perform correlation analysis for all numerical variables.

(10 marks)

Q3 (a) Discuss **THREE (3)** major differences between Ordinary Least Square Method and Logistic Regression using appropriate example.

(9 marks)

(b) **Figure Q3** shows the summary of the generated model of linear regression to predict the output y.

(i) Describe the explanatory power of the selected features towards response variable.

(10 marks)

(ii) Based on the R-Squared value, write your conclusion about the model. Propose the strategy to improve the performance of the model.

(6 marks)

Q4 (a) Business intelligence is a powerful tool in data analytics. It provides solutions in decision making strategies for organisations. Explain **TWO (2)** examples that makes use of the business intelligence concept.

(6 marks)

(b) Vehicle accident data was taken for 8 days in 3 roadways. These roadways are classify as dangerous for road users. During the monitoring duration, the number of accidents were recorded based on two types of vehicle; which are cars and motorcycles. Other observations such as the condition of the weather during the accident were also recorded. The recorded data are presented in **Table Q4**.

(i) Construct a data analytic dashboard by sketching manually the appropriate graphs or figures that shows at least **FIVE (5)** insights that can be extracted from the data.

(10 marks)

(ii) Determine the findings or conclusion that can be obtained based on the dashboard that you have created.

(5 marks)

(iii) Based on your findings in **Q4(b)ii**, suggest suitable solutions to reduce the number of accidents on the 3 roadways.

(4 marks)

– END OF QUESTIONS –

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2020/2021
COURSE NAME : DATA SCIENCE AND APPLICATION
PROGRAMME CODE : BFF
COURSE CODE : BFS 41203

	x1	x2	x3	x4	x5	y
0	10	1	A	0.1	115	Type I
1	12	0	B	0.6	127	Type II
2	18	0	B	0.9	139	Type II
3	17	1	C	0.1	146	Type III
4	15	1	A	0.7	188	Type I

Figure Q2

OLS Regression Results

Dep. Variable:	Y	R-squared:	0.732
Model:	OLS	Adj. R-squared:	0.730
Method:	Least Squares	F-statistic:	387.9
Date:	Thu, 15 Aug 2019	Prob (F-statistic):	2.96e-200
Time:	18:42:03	Log-Likelihood:	-147.90
No. Observations:	716	AIC:	307.8
Df Residuals:	710	BIC:	335.2
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.6957	0.046	-15.250	0.000	-0.785	-0.606
X1	0.1814	0.009	19.220	0.000	0.163	0.200
X2	0.1845	0.009	19.983	0.000	0.166	0.203
X3	0.1702	0.009	18.200	0.000	0.152	0.189
X4	0.1913	0.009	20.961	0.000	0.173	0.209
X5	0.1954	0.009	21.547	0.000	0.178	0.213

Omnibus:	5.569	Durbin-Watson:	1.845
Prob(Omnibus):	0.062	Jarque-Bera (JB):	4.274
Skew:	-0.069	Prob(JB):	0.118
Kurtosis:	2.648	Cond. No.	20.7

Figure Q3

FINAL EXAMINATION

SEMESTER / SESSION : SEM II / 2020/2021 PROGRAMME CODE : BFF
COURSE NAME : DATA SCIENCE AND APPLICATION COURSE CODE : BFS 41203

Table Q4 The energy consumption for the FKAAB building

Date	Road Name	Weather	Car Accident	Motorcycle Accident	Total Accident
1st January	Jalan Bravo 1	Rain	4	7	11
1st January	Jalan Bravo 2	Rain	2	3	5
1st January	Jalan Delta 1	Rain	1	0	1
2nd January	Jalan Bravo 1	Hot	0	4	4
2nd January	Jalan Bravo 2	Hot	1	2	3
2nd January	Jalan Delta 1	Hot	0	0	0
3rd January	Jalan Bravo 1	Slight Rain	3	4	7
3rd January	Jalan Bravo 2	Slight Rain	1	4	5
3rd January	Jalan Delta 1	Slight Rain	1	1	2
4th January	Jalan Bravo 1	Heavy Rain	5	11	16
4th January	Jalan Bravo 2	Heavy Rain	2	6	8
4th January	Jalan Delta 1	Heavy Rain	2	2	4
5th January	Jalan Bravo 1	Rain	2	3	5
5th January	Jalan Bravo 2	Rain	4	3	7
5th January	Jalan Delta 1	Rain	3	3	6
6th January	Jalan Bravo 1	Hot	4	4	8
6th January	Jalan Bravo 2	Hot	1	3	4
6th January	Jalan Delta 1	Hot	1	1	2
7th January	Jalan Bravo 1	Heavy Rain	8	15	23
7th January	Jalan Bravo 2	Heavy Rain	6	6	12
7th January	Jalan Delta 1	Heavy Rain	3	2	5
8th January	Jalan Bravo 1	Slight Rain	4	2	6
8th January	Jalan Bravo 2	Slight Rain	1	5	6
8th January	Jalan Delta 1	Slight Rain	1	6	7