QBUS3600 Assignment 1 (A2B)

Due dates: Thursday 9 April 2020

Value: 30%

Notes to Students

1. The assignment MUST be submitted electronically to Turnitin through QBUS3600 Canvas site. Please do NOT submit a zipped file.
2. The assignment is due at **16:00pm on Thursday, 9 April 2020**. The late penalty for the assignment is 5% of the assigned mark per day, starting after 16:00 on the due date. The closing date **Thursday, 16 April 2020**, 16:00 is the last date on which an assessment will be accepted for marking.
3. Your answers shall be provided as a word-processed report (Microsoft Word, LaTeX or equivalent) giving full explanation and interpretation of any results you obtain. Output without explanation will receive **zero** marks.
4. Be warned that plagiarism between individuals is always obvious to the markers of the assignment and can be easily detected by Turnitin.
5. The data sets for this assignment can be downloaded from Canvas. The dataset is **highly confidential,** and you have responsibility to keep it secure and for it to be used only for your QBUS3600 course work.
6. Presentation of the assignment is part of the assignment. Marks are assigned for clarity of writing and presentation.
7. You should submit your Python code or Jupyter Notebook to the separate submission portal available on canvas
8. You may insert small sections of your code into the report for better interpretation when necessary. However, you must consider the audience of your report.
9. Think about the best and most structured way to present your work, summarise the procedures implemented, support your results/findings and prove the originality of your work.
10. Numbers with decimals should be reported to the **third** decimal point.

Background

Taxi prices are heavily regulated across Australia. In NSW, these regulations are detailed in the "**Point to Point Transport (Fares) Order 2018**" legislation. Fares are the aggregate of five charges: base fare, booking fee, government levy, charge per minute or charge per km (based on a cross-over speed). Customers may also incur extras such as tolls and time-based tariffs. Due to the multi-faceted nature of taxi charges, there are many approaches which can be taken to estimate fares, such as high-level estimation using historical data or more granular methods, such as fare table estimation, including different hourly predictions.

Scenario Story: It is March 2020, and you have just recently started your job as a business analyst at **A2B Australia Limited (A2B),** the parent company to several of Australia's most iconic taxi brands, notably 13Cabs and Silver Service. Your job has been smooth so far and you spend your days sipping jasmine tea and querying SQL databases. One morning as you make your way into the office, breakfast in hand, you spot Ali exiting the Chief Operating Officer's (COO) office, deep in thought. As you attempt to sneak past to your desk, he notices you and walks over.

Ali tells you Fred, the COO, is interested in taxi fare prediction as the general sentiment amongst frequent taxi users is for the addition of upfront payment options for trips, to add certainty for both drivers and passengers. Ali is certain that the in-house analytics team can develop an accurate solution. As most of the department is on holiday, Ali entrusts this important task to you and suggests you start working alone on understanding the data and relevant business.

Initially, Ali is interested in patterns available in taxi fares in Sydney.

Data

The taxi trip data will be provided. This data will include trips in Sydney between August 2019 to October 2019.

BookingID:	Unique identifier for each trip.
M_On_Time:	Records when taxi meter was switched on for the trip. Data is stored in YYYY-MM-DD HH:MM:SS format.
Lat_M_On:	Latitude of start point.
Lat_M_Off:	Longitude of start point.
M_Off_Time:	Records when taxi meter was switched off for the trip. Data is stored in YYYY-MM-DD HH:MM:SS format.
Lat_M_Off:	Latitude of end point.
Lat_M_Off:	Longitude of end point.
City:	Designates the city where transaction took place.
Dest_Suburb:	Records the destination suburb of the passenger.
DelJobDistance:	Records the total distance of the trip. Data is stored in metres.
DelJobTime:	Records the total time of the trip. Data is stored in seconds.
ChargesPrice:	Records the cost of the trip. Data is stored in dollars.
ChargesExtras:	Records any surcharges for trip. Data is stored in dollars.
ChargesFlagfall:	Records the fixed start fee for taxi. Data is stored in dollars.

Task 1 – Preliminary Investigation (80 marks)

Working in a team of analysts at A2B you have been placed in charge of leading the preliminary investigation. This investigation, in the form of a technical report, is intended to serve as a reference for other business analysts working in your team.

You must address the following key questions:

1. Is it possible to establish any relationship among variables or features recorded in the dataset? If so, provide detail about the nature of the relationships.

    For example (these are just examples, you are encouraged to think of other relationships)
    a. Is the charged trip cost relevant to the distances, locations, pick-up times, durations etc?
    b. What kinds of trips are more often?
    c. How is the trip time related to trip distance? The fare regulation provides a formula, but in the real world, traffic, routes, and driving conditions impact fares. What is the relationship between trip distance and fare under different conditions?
    d. How does taxi transport change throughout the week (e.g. Monday compared to Friday)
    e. And more (anything you think worth to explore)

2. How would you segment the trips in the given dataset? What are some segments, and their characteristics? Some examples of ways you could segment trips:
    a. By how much is the trip cost
    b. By how long is the trip distance
    c. By locations where passengers are picked up
    d. By locations where passengers are dropped off
    e. By the extra or sur-charge passengers paid
    f. Can you make some assumptions about the above to translate clusters to passenger or trip type (work, leisure, etc.)
    g. How does Sydney move in taxis? How do pick up and drop off areas change throughout the day, or the week?

To address **Questions 1 and 2,** you are expected to perform a preliminary Exploratory Data Analysis (EDA) over the dataset. You are expected to find or reveal all possible properties, characteristics, patterns and statistics hidden in the dataset with a view to the problem at hand.

You are limited to a maximum of 15 pages. Your report should describe, explain, and justify your findings. Ensure that your report is concise and objective.

List key resources as references in the end of your report, such as journal articles, conference papers, reports, news and software etc. Use APA style for your references.

Task 2 – Executive Briefing (20 marks)

You have been asked to summarise your findings so that they can be shared with the wider business and in particular, management. This one-page briefing should concisely describe your findings to a non-technical audience and primarily address the business problem. In the briefing you should also outline your suggestions for acting on your findings.

You are limited to a maximum of 1 page.

Marking and Key Rules

Your reports will be marked against the following criteria and rules:

- Demonstrate a clear understanding of the problem
- Demonstrate consideration for the audience
- Clear outline and demonstration of investigation process
- The analysis overall is sound and logical
- Clearly draw conclusions based on analysis
- Statements are clear, concise and accurate, with correct spelling, free of grammar errors and correct use of punctuation
- Use of visual presentation is appropriate
- The report is well structured, and sentences are well connected
- Closely follow a referencing style specified in Business School Referencing Guide (e.g. APA) with consistency
- Clear, concise and commented Python code, if any