Mock Exam

Social Media and Web Analytics

Spring 2024

Instructions to Candidates

The exam is graded out of **100 points**. Points for each of the three parts are clearly labelled at the beginning of each. The point allocation for each question is also clearly designated.

The **exam duration is 180 mins**.

Write your Student ID and name on <u>every</u> page of the exam.

Answer each question in the text box directly underneath the question. There are additional blank pages at the end of the exam if you need to cross out your work and start again. Clearly label your answers on any of the extra pages.

Answers to the questions should be **based on the course material presented in lectures, Lab Sections and readings.**

Remember, **your goal is to communicate**. Full credit will be given only to the correct solution which is described clearly. **Convoluted and obtuse descriptions might receive low marks, even when they are correct**. Also, aim for concise solutions, as it will save you time and also help you conceptualise the key idea of the problem.

No calculators or dictionaries are permitted.

Any **communication between students during the exam**, no matter whether it is about the exam or not, **is unacceptable** and will likely result in harsh penalties. Attempting to view other student's answers will be treated in a similar manner.

To pass Social Media and Web Analytics, you must receive a grade higher than 5.5/10 for this exam.

Part A: Multiple Choice (30 points)

Answer *all* questions. **Each question in Part A is worth 3 points**. There is **no** guessing correction applied to the grading.

Question 1.
Which statement best describes a key difference between the BING and VADER sentiment lexicons?

A. BING provides sentiment scores for individual words, while VADER provides a compound sentiment score for entire texts.

B. BING is optimized for analyzing social media text, while VADER is designed for general-purpose sentiment analysis.

C. BING does not account for sentiment intensity, while VADER accounts for intensity using heuristics and punctuation.

D. BING incorporates contextual understanding of phrases, while VADER relies solely on individual word sentiment scores.

Question 2.
A company implements an A/B test with treatment assignment at the user level and intends to analyse data at the purchase level so that there might be multiple purchases per individual during the experiment. When analysing the data via linear regression analysis, what assumption should they make on the variance of the error term?

A. It is homoskedastic

B. It is auto-correlated over time

C. It is clustered at the user level

D. It is heteroskedastic

Question 3.
Consider the following regression equation:

$$y_{it} = \beta_0 + \beta_1 Treatment_i + \beta_2 After_t + \beta_3 Treatment_i \times After_t + \varepsilon_{it}$$

Assuming the parallel trends assumption holds, the Difference in Differences estimate of the average treatment effect is

A. $\beta_1 + \beta_3$

B. β_3

C. $\beta_3 - \beta_2$

D. $\beta_3 - \beta_1$

Question 4.

Fill in the blanks. When testing motivations for posting in social media, researchers define Intrinsic Utility as _____(i)_____ and predict that experimentally increasing the number of followers would _____(ii)_____ if this was the main driver of posting behaviour.

A. (i) inherent satisfaction of posting; (ii) increase the number of user posts

B. (i) inherent satisfaction of posting; (ii) decrease the number of user posts

C. (i) posting motivated by the perceptions of others; (ii) increase the number of user posts

D. (i) posting motivated by the perceptions of others; (ii) decrease the number of user posts

Question 5.

Under what conditions are the CUPED and Difference in Difference estimate of an average treatment effect from a field experiment the same?

A. Never, these methods cannot be used to analyse the same kind of data.

B. There is **no** differences in the average pre-treatment outcomes *and* $\theta = 1$

C. There *are* differences in the average pre-treatment outcomes *but* $\theta < 1$

D. There **no** differences in the average pre-treatment outcomes *and* $\theta < 1$

Question 6.
A company uses observational data paired with linear regression to analyze the effect of ad exposure on purchase behaviour over a one month time period. Their findings reveal that users who see the ads are more likely to purchase the product. What is the primary reason why this linear regression should not be interpreted causally?

A.   The model does not account for the seasonal variations in sales data.

B.   The sample size is too small to draw meaningful conclusions.

C.   Selection bias, because users who see the ads may differ systematically from those who do not.

D.   The regression assumes a linear relationship between ad exposure and purchase behaviour, which might not be accurate.

Question 7.
What is the primary purpose of word stemming and word lemmatization when analyzing text data?

A.   To remove stop words from text data to reduce dimensionality.

B. To convert words to their root or base form for standardizing text data.

C.   To identify named entities within the text, such as names of people or places.

D.   To detect the sentiment of a given text by analyzing the intensity of words.

Question 8.
In quantitative marketing research, what does the process of **identification** refer to?

A.   The technique of selecting the appropriate model specification for analysis.

B.   The method of gathering a representative sample to ensure generalizability.

C.   The approach of differentiating between endogenous and exogenous variables.

D.   The process of figuring out what part of the variation in your data answers the research question.

Question 9.

What is a document-term matrix in the context of topic modeling?

A. A matrix that represents the frequency of co-occurrence of terms within the same document.

B. A matrix that lists documents as rows and terms as columns, with each cell indicating the frequency of a term in a document.

C. A matrix used to map documents to predefined topics based on their content.

D. A matrix that records the similarity scores between different documents based on their term usage.

Question 10.

Which statement best describes the parallel trends assumption in the context of Difference-in-Differences (DiD) analysis?

A. The assumption that the outcome variable follows the same time trend in both treatment and control groups in the absence of treatment.

B. The assumption that the treatment and control groups have the same number of observations.

C. The assumption that the baseline characteristics of the treatment and control groups are identical.

D. The assumption that the treatment effect is constant over time.

Part B: Short Answer Questions (40 points)

Answer *all* questions in this section. Point allocations per question are clearly indicated by each question and/or sub-question.

Question 1 (10 points). Experiments and Causal Effects.

When discussing this class with your friends who have not taken classes in marketing analytics, one proclaims

"I do not understand why your Social Media and Analytics class focuses on randomised- and natural- experiments to estimate causal effects, regression coefficients from linear regression with observational data delivers causal results."

Do you agree or disagree with this comment? Explain why (max. 5 sentences)

Question 2 (10 points). Online Word of Mouth and Product Demand.

(a) [4 points] List and provide one sentence explanations of two mechanisms through which online word of mouth can influence product demand.

(b) [7 points] Explain which mechanisms that connect word of mouth to consumer demand are important for movies with either small or medium size expected box office.

HINT: Highlight why any differences you identify might occur.

Question 3 (10 points). Donors Sharing About Their Charitable Contributions.

(a) [4 points] Explain the tradeoff that donors face when deciding to share about their charitable donations

(b) [6 points] Explain how researchers tested a new strategy to encourage greater sharing and summarize their findings.

Question 4 (10 points). Topic Modelling.

(a) [4 points]. Assume that you have a data set of Trip Advisor reviews about hotels that has already been cleaned and converted to a Document-Term-Matrix called `reviews_dtm`.

Write R code that estimates a structural topic model with 7 topics that would output the same output if run more than once.

HINT: Assume all required R libraries are loaded.

(b) [6 points] Explain how a marketing analytics team that works for Hilton Hotels could use sentiment analysis together with topic modelling to propose new managerial strategies that improve the consumer experience.

Part C: Structured Essay Question (40 points)

Answer **all** questions.

Scenario:

In your position as marketing scientist at Meta, you and your team are tasked with understanding the importance of having news articles appear in user's Facebook feeds for driving their engagement with *non-news* content and user's content generation. Facebook blocked all news content from appearing in user feeds for 2 weeks in Australia in response to a regulatory decision that was specific only to the Australian market.

Question 1 (2 points).
Explain why news content is an important driver of user engagement on Meta. (max 2 sentences)

Question 2 (2 points).
State the main research question(s) that your team needs to answer.

Question 3 (6 points).

List and define three outcome metrics you will want to measure the effect of the news block. Use simple, yet precise terminology.

Question 4 (6 points).
Write down a research design that allows you to measure the effects of the news block on the outcome variables list above. (Max 8 sentences)
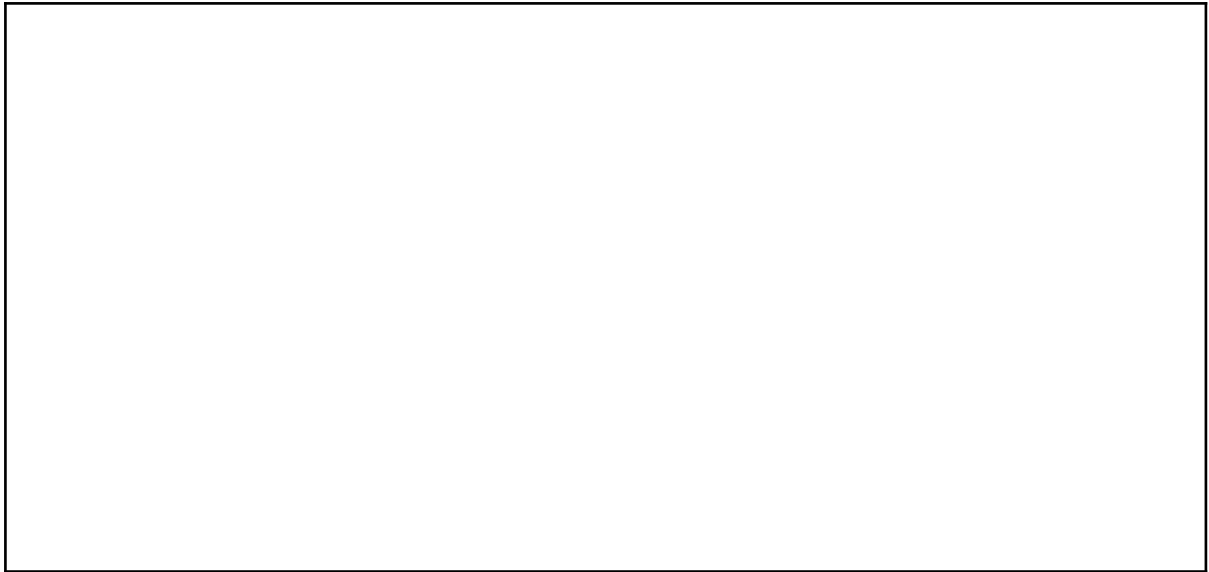
Question 5 (6 points).
Outline the empirical framework you will use, specifying any regression equations or other statistical techniques you will need to answer the question. (Max 8 sentences)
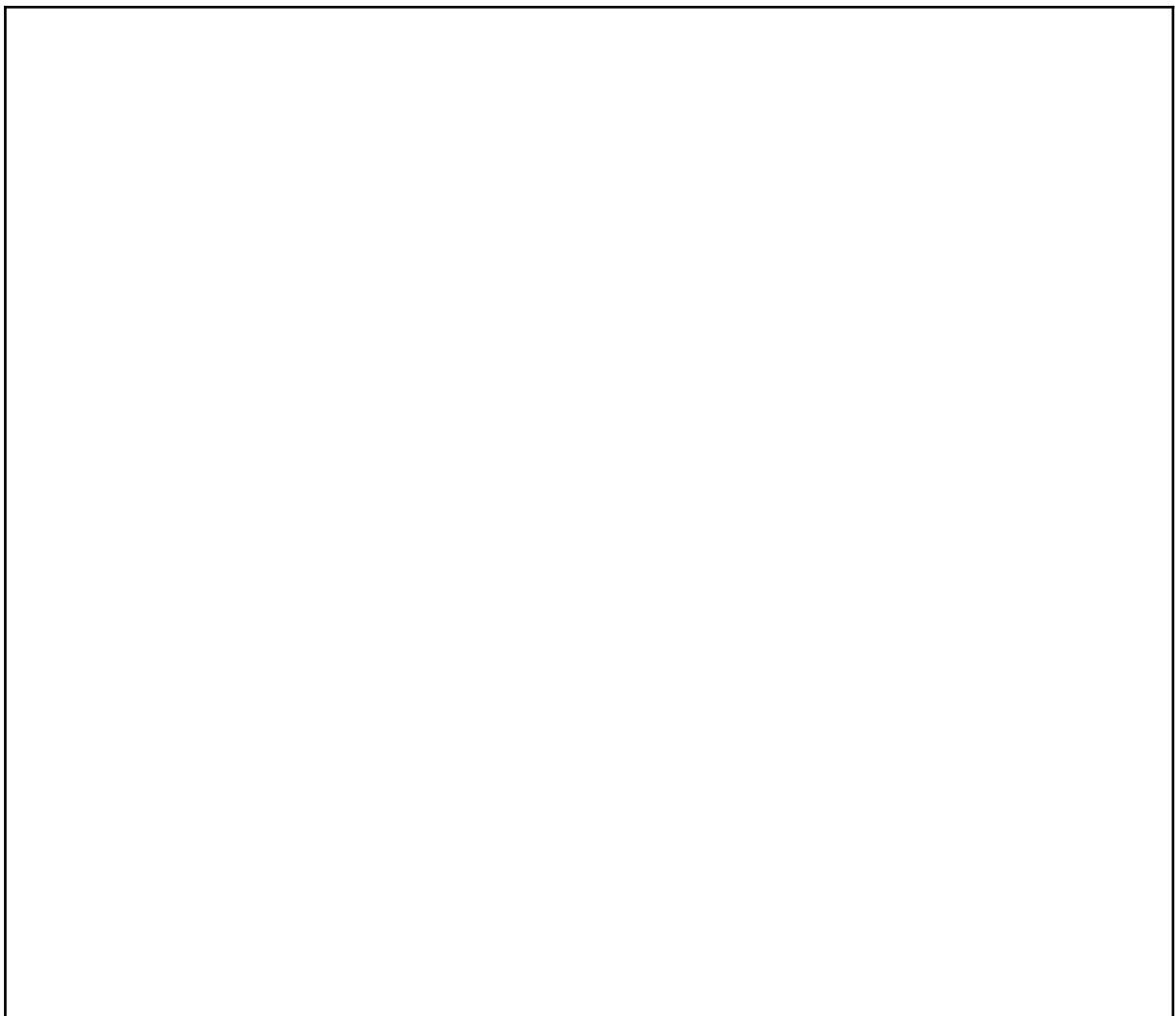
Question 6 (4 points).
How does the empirical modelling you propose deliver answers to the research question(s) outlined in (2)? (Max 5 sentences)
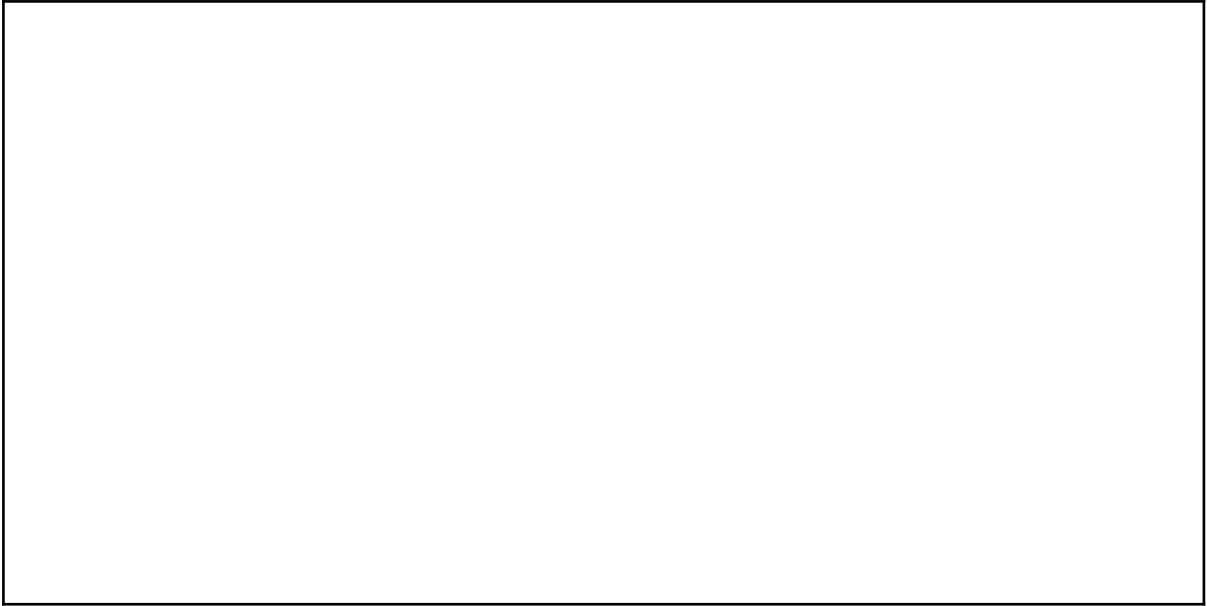
Question 7 (4 points).

What results would you expect to find? Link these back to explicit managerial strategies for Facebook to pursue in the future. (Max 5 sentences)

(END OF EXAM.)