860 Exam July 2024

**General Instructions:**

1. **This exam must be completed in your own original words** – you may not copy / paste or otherwise reproduce text from any source other than Python output and code.  The prohibition on copying material extends to any material that you may have written or assembled prior to the exam (other than code). Copy / pasting material may constitute a departure from academic integrity.
2. The use of Generative AI tools such as ChatGPT are permitted for code generation and troubleshooting, but are not to be used to generate any written responses. Please note that these tools are not always reliable. They will sometimes produce code that does not run, or code that does not do what you expect it to.
3. This exam is a confidential, individual exam. The content of the exam may not be discussed with anyone.  Failure to abide by these rules may constitute a departure from academic integrity.
4. Data, where required, will be found in the appropriate exam data file which will be shared with this document.
5. All analysis and data manipulations should be completed in Python. Submissions can be in Jupyter or Word/PDF form, but only one document will be graded.
6. Generally, you should include in your submission code and output to document your model or analysis where it is called for.  We will not be looking at a separate set of Python code as part of the grading unless we deem it necessary – just the output.
7. Answer both questions. The exam should take less than 2 hours to complete. You will have the full exam period.

Please save your work often with version control. This will reduce the risk and consequence of a possible technology failure. Extensions will not be provided.

* + 1. A screenshot of a computer program

       Description automatically generated
    2. We then started to bui
  1. How do you know this is the best model? Bee specific. **(2 marks)**
  2. Does temperature have an impact on the number of bees? Answer this question using a formal test. State the null hypothesis, the p-value, and your conclusion. **(3 marks)**
  3. If the total number of bees ever goes beyond 95,000, the hive will [swarm](https://en.wikipedia.org/wiki/Swarming_(honey_bee)). Yes, this is a real thing, and yes, it is terrifying. Not to mention, you often lose your bees! The good news is, if you can predict a swarm, you can give them more room and prevent it. Use the data on the tab ‘NextSeason’ to predict the size of the beehive in each month of the next season. The data is sourced directly from the Farmer’s Almanac. Should I be worried about swarming if the forecasted weather is correct? Why or why not? **(4 marks)**

1. You are interviewing for a job at a prestigious consulting firm that focuses on solving business problems using analytics. The following are the questions asked by the recruiting manager – they are based on real client problems and questions. For each question, provide a clear and concise answer that shows you understand the material from **both** a managerial and statistical perspective.
   1. Your client is proposing a satisfaction survey which should generate about 5000 responses. One of the questions has potential responses ‘very dissatisfied’, ‘dissatisfied’, ‘no opinion’, ‘satisfied’ and ‘very satisfied’. Since there is a natural ordering to these results, he is not sure if he should code the results 1,2,3,4,5 respectively or -2,-1,0,1,2 respectively. Explain and justify your thoughts on how the data should be coded, and any strengths and weaknesses the choice of coding has. **3 marks**
   2. How could you find a multi-dimensional outlier in a linear regression context? What should you do about it? **3 marks**
   3. Your client is worried about the restrictions imposed on a linear model. She suspects some of the relationships have non-linear patterns like decreasing returns to scale or an exponential shape. What would you tell her? **3 marks**