
CLARITY
INSIGHTS
Data Science Challenge

Instructions

Purpose	This case study is an opportunity for you to demonstrate your ability to • Perform basic data processing • Identify appropriate analytics techniques / models to address a typical business question • Draw insights from data • Communicate these insights to a business audience
Instructions	Please create a presentation that addresses the question posed in the case description • You may use any tool (e.g. R, python) to perform the analysis but please indicate which tool you used within the presentation • If needed, you may use any machine learning, AI or statistical technique which you consider to be appropriate • The case study includes a basic set of data that can/should be used for the analysis, however feel free to supplement this data with any openly available data that you choose. If you use supplemental data, please include the source of the data as a footnote. • Clearly indicate any assumptions which you are making • You will have 1 hour for your presentation, including at least 15 minutes for Q&A. • Assume you will be presenting to an audience primarily concerned with the business outcomes / insights you have discovered and less concerned with a detailed technical explanation of the techniques you have used. If a deep technical explanation is needed to fully understand your work, then include it as backup material. • Prior to your interview, please email your your solution as PowerPoint, Google Slides, or pdf.
Judging Criteria	Your submission will be judged on your ability to • Map an ambiguously defined business challenge into a problem addressable via analysis and data science • Identify appropriate techniques (including exploratory analysis, machine learning, etc.) for performing this analysis • Ability to interpret the results of the modelling (what do the coefficients mean, how do you use them?) • Extract non-trivial insights, conclusions and recommendations from data based upon these analyses • Communicate these insights, conclusions and recommendations to a business audience that is not comprised of experts in data science

CLARITY INSIGHTS

Sales Forecasting

Business Problem:
Build and evaluate models to predict national retail store sales for each store and department. Our sales are very seasonal and we make our money during holidays like the Super Bowl, Labor Day, Thanksgiving, and Christmas.

Demonstrate your model, by creating forecasts for the following periods:

- Month of November 2012
- Total forecast sales for Q4 of 2012 (where Q4 includes October, November, and December of 2012)
- Total forecast sales for Q1 of 2013 (where Q1 includes January, February and March of 2013)

Please be sure to explain how you performed these predictions and what (if any) limitations or risks there may be in the predicted values

Data Sources:
- Sales by Store & Department [CSV]

Common questions

1. Is there a data dictionary or some background on the data set/variable? If not, I'm happy to proceed making certain assumptions, which I will state in the deck.

2. Unfortunately, no additional information is available for this data set. I would suggest that you do as proposed i.e. make whatever assumptions you think are needed and then document them.


Good Luck