Social Finance Interview Data Analytics Exercise

Measuring recidivism rates

DATA ANALYTICS EXERCISE OVERVIEW:

Social Finance's analytics interview is designed to test your data analysis and visualization skills, as well as your ability to synthesize and communicate results. We are most interested in understanding your thought process and approach to analytical problems. In an effort to make our interview time as efficient as possible, we ask that the data analytics exercise be completed in advance. It should take approximately 3 to 4 hours to complete, and you will have a chance to talk through your work as part of the in-person interview. The completed exercise is required to be sent back to Social Finance **no later than 24 hours before the scheduled in-person interview** (please refer to your informational email regarding specific due date timing).

BACKGROUND: As part of a project, you are working on studying the effects of a program to reduce recidivism (which is defined as returns to incarceration). To understand the baseline landscape, you have been asked to calculate historic recidivism rates using data from the client, the State Department of Corrections (DOC). An analyst from DOC has provided Social Finance with a series of raw data files representing individuals released from the state prison system in each year from 2010 to 2014. Each file contains three fields:

- ***pers_id***, a unique identifier that represents an individual person. pers_id is consistent across files, so every instance of the pers_id field across all years represents the same person.
- ***sentence_date***, which represents the date that an individual was originally sentenced. For the purpose of this exercise, you can assume that date represents their first day of incarceration.
- ***release_date***, which represents the date the individual was released from incarceration.
- ***crime_type,*** which represents the category of crime for which the individual was incarcerated.
- ***education_level,*** which represents the highest level of education an individual has achieved.

EXERCISE INSTRUCTIONS:

Your task is to calculate the 1-, 2-, and 3-year recidivism rates for individuals first released from incarceration in 2010. The goal is to be able to answer the question **"what percentage of individuals released in 2010 were incarcerated again within 1 year, within 2 years, and within 3 years?"**. For this analysis, you can assume that we are only interested in the first instance of re-incarceration an individual may experience – so an individual who was re-incarcerated within one year of their original release will by definition have also been re-incarcerated within three years of their original release.

In addition to the variables provided in the original file, DOC has provided an additional file which contains risk scores calculated for individuals. Only "medium" and "high" risk offenders are subject to incarceration; "low" risk offenders are sentenced to probation. **How do the recidivism rates you originally calculated vary by risk level?** Feel free to also explore the crime type and education variables if you find that incorporating them into your

analysis is compelling.

Note: As is true of almost all raw data, the data shared by the Department of Corrections includes data quality issues. You'll need to address these issues the best way that you see fit before conducting your analysis. We will be interested to discuss how you think through these challenges and the implications for your analysis.

1. As part of your submission, please provide:
    a. Any code you developed to conduct the analysis
    b. A compelling visual representation of your findings
    c. A list of bullet points summarizing the key takeaways
    d. A 3-4 paragraph narrative of your findings, including in particular:
        i. Understanding the limitations of the available data, what preliminary takeaways would you bring back to the Department of Corrections?
        ii. A brief description of why you chose to approach the analysis in the way you did
        iii. A brief speculation on what other data fields you would like to have that might allow you to perform a more sophisticated analysis or provide a better understanding of recidivism patterns, and how you would incorporate them into your analysis.
        iv. Your thoughts on the data equity implications of your analysis.
        v. Any other thoughts or findings you would want to elevate as part of this analysis.
2. If you have questions as you complete this case, we encourage you to make an assumption and clearly document it.
3. Once completed, please send your analysis and output back to Social Finance in whatever format you feel is most appropriate. Please refer to the email in which you received these materials for your point of contact. As a reminder, these files should be sent back no later than 24 hours before your scheduled in-person interview.