HW Data Visualization with Python

Background

This HW is designed to help you learn data visualization using Python packages (e.g., matplotlib, seaborn). Use the dataset (bank.csv). **Your client is a German bank, lending money to individuals. They want to find out:**

- **The characteristics of the people who pay back**
- **The characteristics of the people who default on their loans.**

- German Credit Dataset (Real-world dataset) https://archive.ics.uci.edu/ml/datasets/Statlog+%28German+Credit+Data%29
- Description of dataset http://iainpardoe.com/teaching/dsc433/handouts/German.pdf

1. OBS# Observation No. Categorical
2. CHK_ACCT Checking account status Categorical
    0 : < 0 DM
    1: 0 < ...< 200 DM
    2 : => 200 DM
    3: unknown
3. DURATION Duration of credit in months Numerical
4. HISTORY Credit history Categorical
    0: no credits taken
    1: all credits at this bank paid back duly
    2: existing credits paid back duly till now
    3: delay in paying off in the past
    4: critical account
5. NEW_CAR Purpose of credit Binary car (new) 0: No, 1: Yes
6. USED_CAR Purpose of credit Binary car (used) 0: No, 1: Yes
7. FURNITURE Purpose of credit Binary furniture/equipment 0: No, 1: Yes
8. RADIO/TV Purpose of credit Binary radio/television 0: No, 1: Yes
9. EDUCATION Purpose of credit Binary education 0: No, 1: Yes
10. RETRAINING Purpose of credit Binary retraining 0: No, 1: Yes
11. AMOUNT Credit amount Numerical
12. SAV_ACCT Average balance in savings account Categorical
    0 : < 100 DM
    1 : 100<= ... < 500 DM
    2 : 500<= ... < 1000 DM
    3 : =>1000 DM
    4 : unknown
13. EMPLOYMENT Present employment since Categorical
    0 : unemployed
    1: < 1 year
    2 : 1 <= ... < 4 years
    3 : 4 <=... < 7 years
    4 : >= 7 years
14. INSTALL_RATE Installment rate as % of disposable income Numerical
15. MALE_DIV Applicant is male and divorced Binary 0: No, 1: Yes
16. MALE_SINGLE Applicant is male and single Binary 0: No, 1: Yes
17. MALE_MAR Applicant is male and married or widower Binary 0: No, 1: Yes

18. CO-APPLICANT Application has a co-applicant Binary 0: No, 1: Yes
19. GUARANTOR Applicant has a guarantor Binary 0: No, 1: Yes
20. TIME_RES Present resident since - years Categorical
    0: <= 1 year
    1<…<=2 years
    2<…<=3 years
    3:>4years
21. REAL_ESTATE Applicant owns real estate Binary 0: No, 1: Yes
22. PROP_NONE Applicant owns no property (or unknown) Binary 0: No, 1: Yes
23. AGE Age in years Numerical
24. OTHER_INSTALL Applicant has other installment plan credit Binary 0: No, 1: Yes
25. RENT Applicant rents Binary 0: No, 1: Yes
26. OWN_RES Applicant owns residence Binary 0: No, 1: Yes
27. NUM_CREDITS Number of existing credits at this bank Numerical
28. JOB Nature of job Categorical
    0 : unemployed/ unskilled - non-resident
    1 : unskilled - resident
    2 : skilled employee / official
    3 : management/ self-employed/highly qualified employee/ officer
29. NUM_DEPEND Number of dependents Numerical
30. TELEPHONE Applicant has phone in his or her name Binary 0: No, 1: Yes
31. FOREIGN Foreign worker Binary 0: No, 1: Yes
32 RESPONSE Fulfilled terms of credit agreement Binary 0: No, 1: Yes

- **Column #32 ('RESPONSE')** is Y value in regression, meaning that is something you're trying to predict. Therefore, your analysis should focus on who is likely to "fulfill" the credit agreement. 1 means they DID and 0 means they DID NOT.

Requirements

- Rename **HW_DataVisualizationwithPython_s.ipynb** as "**yourlastname_firstinitial_ HW_DataVisualizationwithPython_s.ipynb**") and complete all the **tasks** in the notebook. **Each cell must be properly numbered and formatted using Markdown**.
- For any questions regarding Python coding, you should consult lecture notes (Jupyter Notebooks) and **https://stackoverflow.com/**
- Your Jupyter Notebook needs to be properly formatted using Markdown and comment.
- For each chart, you must include xlabel, ylabel, and title (if available).

Questions:

1. Use some common sense (and/or your business knowledge) to answer the first question. Which columns or (independent) variables would significantly influence the Y value (whether someone fulfills the terms of credit agreement or not)? List at least three and explain why?

2. What are some general findings from basic statistics (describe)?

3. What portion of borrowers have paid back? What portion have defaulted on their loans?

4. What is the relationship between RESPONSE and three other variables (Choose three variables for this analysis based on your answer for Question#1)?

5. What are some insights from the chart about the relationship between DURATION and RESPONSE?

6. What variables appear to be highly influential in determining Y value (RESPONSE)?
7. Be very specific. What are the characteristics of the people who have paid back? Be very specific. What are the characteristics of the people who have defaulted on loans?
8. Provide additional insights from interactive plots using Bokeh.
9. Provide additional insights from interactive plots using plotly.
10. What is your general recommendation for the German bank?

How to Format Your Jupyter Notebook:

- Start with K-State Honor Code "**On my honor, as a student, I have neither given nor received unauthorized aid on this academic work**."
- Include the question and the question number, using Markdown, prior to each cell containing python codes.
- Must be professional and neat (You are submitting this report for consideration by your upper managers

Submission

- Complete Ipython notebook in **HTML** version. Download as HTML. (yourlastname_firstinitial_HW_DataVisualizationwithPython.**html**)