Retro Rabbit Data Scientist Test

Hello potential Retro Rabbit Data Scientist!

This test aims to evaluate the following abilities that a Retro Rabbit data scientist requires:

- Exploratory data analysis
- Data scrubbing
- Understanding different modelling techniques and performance measures
- Communication
- Data science process

Please note the following

- The test should take between **4 - 8 hours** depending on level of experience
- You answer should be submitted in the form of a [Python (jupyter) notebook](#) or [R notebook](#) (or two) as an email attachment or github repo to ja@retrorabbit.co.za
- **Answers will be evaluated for plaigerism**
- You have from the *1st of November* until the **6th of November** to complete the test
- Please direct all questions to Jade Abbott: ja@retrorabbit.co.za. She will do her best to answer within 24 hours of any question.

Task 1: Exploratory Data Analysis

Download the following dataset: https://www.kaggle.com/ludobenistant/hr-analytics

In your notebook, perform the following tasks:
1. Scrub the data
2. Perform exploratory data analysis that will provide insight into the data
3. Summarize your findings

Task 2: Modelling

Download the following dataset: https://www.kaggle.com/dalpozz/creditcardfraud
You are tasked to predict whether someone would perform credit card fraud

1. Preprocess the data by performing appropriate data cleaning, transformations and feature selection
2. Suggest a technique to predict the target feature. Please write your answer in the notebook, and write why you suggest that model. Also please select appropriate performance measures
3. Run your selected technique and report on your results from using the technique
4. Given your results, suggest the next steps on how to further improve the model

Good luck!

Kind Regards,
Retro Rabbit Team