Data Analysis with Python

Project Brief

Project Deliverable

- Your deliverable will be a python notebook that will contain your solution.
- You will need to submit the shareable link to your notebook

Problem Statement

While applying for university, foreign student populations could greatly benefit from data and resources to support their wellbeing and success. Such students and families often lack the necessary information to distinguish between their school options, access services, and identify affordable housing near the high-quality school and in safe neighborhoods with access to transit and employment.

Jane is a 20-year-old high school graduate from Nigeria. She has recently completed her high school education and has decided to pursue a degree in Management Systems and Information Technology in the United States.

Jane has approached your university recruiting agency and has tasked you to help her search for the best school for her. She is willing to relocate anywhere in the continental United States, but she has a few criteria that her excellent schools must satisfy:

- safety (low crime),
- urban -- Jane wants to live the big city life, and
- start-ups -- the school should be in a metropolitan area that ranks highly in entrepreneurialism (she plans to find an internship at a startup while she studies).

Jane would like you to help her narrow down her search to a list of schools that she can investigate more closely before deciding. You need to produce a dataset of schools that satisfy all of Jane's criteria, ranking them from best to worst according to the same criteria.

Jane's schools must:

- Be in an urban/metropolitan area.
- Be in a city that ranks 75th percentile or higher on Kauffman's start-up rankings.
- Be below 50th percentile in overall crime.
- Offer a 2-year or 4-year degree in Information Technology/Science.

Dataset Download Link: https://bit.ly/2ZiWu9P

Dataset Download Link II: https://bit.ly/2S1n03e

Hints:

- Read the data dictionaries to figure out what the variables mean and which ones you will need to use.
- Eliminate unneeded columns.
- Look for suitable columns to join the tables.
- Perform any cleaning and standardization needed to facilitate the joins.
- Engineer a summary variable for school crime so that we can compare schools by levels of crime overall.
- Eliminate from the data all the data points that fail to satisfy Jane's criteria.
- Engineer a method for ranking the schools in consideration of all of Jane's criteria taken together.

You can use the following guiding notebook to get started.

Source: [https://data.world/opportunity]