

1

Project 1: Standardized Testing, Statistical Summaries and Inference

2

3

4

Overview

5

6

Our first module in DSI covers:

7

– basic statistics (distributions, confidence intervals, hypothesis testing)

8

– many Python programming concepts

9

– programmatically interacting with files and directories

10

– visualizations

11

– EDA

12

– working with Jupyter notebooks for development and reporting

13

14

You might wonder if you're ready to start doing data science. While you still have ****tons**** to learn, there are many aspects of the data science process that you're ready to tackle. Project 1 aims to allow you to practice and demonstrate these skills.

15

16

For our first project, we're going to take a look at aggregate SAT and ACT scores and participation rates from each state in the United States. We'll seek to identify trends in the data and combine our data analysis with outside research to identify likely factors influencing participation rates and scores in various states.

17

18

Generally speaking, you will be asked to come up with a data science problem. Here's a specific prompt that should help you craft this statement:

19

> The new format for the SAT was released in March 2016. As an employee of the College Board – the organization that administers the SAT – you are a part of a team that tracks statewide participation and recommends where money is best spent to improve SAT participation rates. Your presentation and report should be geared toward ****non-technical**** executives with the College Board and you will use the provided data and outside research to make recommendations about how the College Board might work to increase the participation rate in a ****state of your choice****.

20

21

22

23

Datasets

24

25

Provided Data

26