

DS 4002

Data Science Project Course

First Session

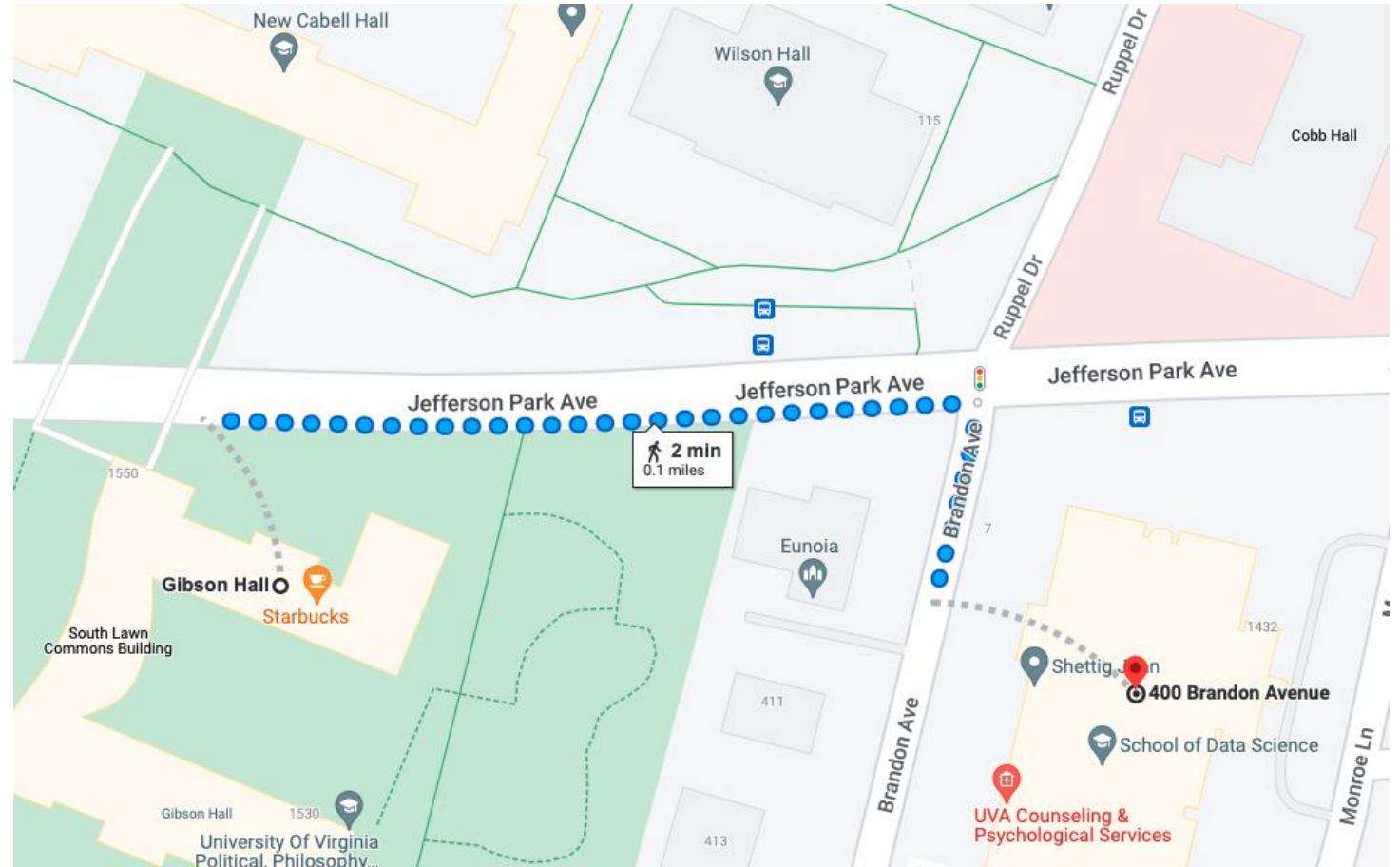
2022-08-24

Warm Up – Collect Data

- Where did you grow up?
- What is your favorite beverage?
- Why did you enroll at UVA?

Instructor introduction

- Professor Alonzi
 - lpa2a@virginia.edu
 - Elson 185A
- Harsh Anand
 - yyf8rj@virginia.edu
 - Link Lab



Learning Objectives

- Cognitive
 - synthesize relevant ideas in a domain and create a testable hypothesis
 - translate processes from theory in the scientific method into functioning data science pipelines (technical focus)
 - develop experimental plan based on the scientific method principles
- Social
 - demonstrate appropriate and constructive behavior when providing critical feedback
 - coordinate effort with teammates to accomplish goals within a given set of constraints
- Psychomotor
 - Present results of a team project and field audience questions

Method

- Schedule
 - First two weeks (form) – walk through course project framework
 - 3 x four week cycles (reps) – execute framework on 3 projects
- Group Work
 - Groups of 3 for all three projects
 - Designated different group leader for each project
- Specifications grading (more later)

Collect Data

- What is your pump-up jam?
- Is a hot dog a sandwich?
- What is one goal you have for this course? ... and why?

Specifications Grading

- This system is very different from the mean-weighted average system
 - Every assignment has a rubric
 - Every assignment is graded pass/fail
 - Jargon “meets spec” or “does not meet spec”
 - Resubmissions are permitted
 - Letter grades are earned by completing bundles of assignments

Assignments	C Bundle	B Bundle	A Bundle
(MI 1) Create Hypothesis	3	3	3
(MI 2) Exploratory Analysis/Pipeline Plan/Set Spec	3	3	3
(MI 3) Analysis/Pipeline Construct	3	3	3
(MI 4) Prepare Results for presentation	3	3	3
(PR 1) Present Results (Oral)	1	1	1
(CC 1) Constructive Critique		1	2
(IR 1) Intangible Reflection Slide	1	2	3
(CS 1) Case-study: review & reflection	1	2	3
(CS 2) Case-study: produce (manuscript)			1

Example Rubric – “MI1 Create Hypothesis”

Submissions that meet spec:

- One-page pdf containing a hypothesis sentence and justification paragraph
- Synthesize relevant ideas in a domain and present a testable hypothesis
- Include an explicit hypothesis sentence
 - One Sentence
 - Quantifiable
 - Testable
 - E.G. meets spec: Michael Jordan lead the NBA in scoring in 1992-1993 season.
 - E.G. does not meet spec: Michael Jordan was the best player in the 1992-1993 NBA season.
- Include a researched justification paragraph
 - Contains Motivation: Why is this hypothesis worth testing?
 - Contains Current State of Knowledge: What is currently known?
 - Contains References

Schedule

Week	Date	Project Number	Project Cycle	Plan	Milestone
1	August 24, 2022	0		Course Philosophy	
2	August 31, 2022	0		DS project fundamentals	
3	September 7, 2022	1	1	Begin project 1	MI 1
4	September 14, 2022	1	2		MI 2
5	September 21, 2022	1	3		MI 3
6	September 28, 2022	1	4	Presentation Creation and Delivery	MI 4
7	October 5, 2022	2	1	Begin project 2	MI 1
8	October 12, 2022	2	2		MI 2
9	October 19, 2022	2	3		MI 3
10	October 26, 2022	2	4	Presentation Creation and Delivery	MI 4
11	November 2, 2022	3	1	Begin project 3	MI 1
12	November 9, 2022	3	2		MI 2
13	November 16, 2022	3	3		MI 3
14	November 23, 2022	Thanksgiving Recess -- No class			
15	November 30, 2022	3	4	Presentation Creation and Delivery	MI 4
Exam Period	December 14, 2022	No exam. Last day to submit assignments (e.g.: the A-level case study or project 3 resubmissions)			

Collect Data

- What is your favorite programming language?
- What is your favorite probability distribution?
- What is your favorite cheese?

Framework

1. Research Topic
 2. Create Hypothesis
 3. Acquire/Establish Data Set
 4. Exploratory Data Analysis (EDA)
 5. Create Analysis Plan
 6. Conduct Analysis
 7. Evaluate Result, Eureka?
 1. If no: goto 5 (or 3)
 2. If yes: goto A
- A. Switch Gears to Presentation
 - B. Create Presentation
 - C. Deliver Presentation

First Two Weeks – Framework Walkthrough

- Professor Alonzi is the designated leader

Pump up Jam time

Step One: Research Topic - Sandwiches

- What is the telos of a sandwich?
- Open question: Is a hot dog a sandwich?
- Pop Culture: [Yahoo Answers](#), [Today Show](#), [r/changemyview](#)
- Oscar Meyer (yes): [official twitter](#)
- NHDSC (no): [Official Policy](#) and "[Argument](#)"
- Merriam-Webster (yes): [Official Definition](#)
- Data: [Twitter poll](#) - 2,022 people - 81.9% no / 19.1% yes - from 2018
- Data: [The Takeout](#) - 34 pop culture figures - 55.9% no / 26.5% yes

Step Two: Create Hypothesis

- Starting Point: Half of UVA students are wrong about whether a hot dog is a sandwich?
- << Class discussion about what our hypothesis should be >>

Step Three: Acquire/Establish Data

- We have some data from class today but we need more
- Setup our database
- Establish our data dictionary
- Collect Data (homework)
- Steps 4-10 next time