Introduction:

[0] Opening Slide

- ## The search is over - It's closer than you think

[1] Original Problem: Track Spikes in San Francisco

- ## I joined club track team, was back in SF

I joined the club track team in December, and then was at home in San Francisco for Winter Break. I needed to buy track spikes the meet that was happening when I returned to campus

- ## Unique Product: Try on & Not Running shoe

Unique product because:
- Needed to try on, so couldn't just get them shipped
- Different from "running shoes"

- ## Idea: Finding products nearby

- I knew the small local sports stores from playing high school sports, but they didn't come up when I was searching for 'track spikes'

[2] My Independent Study: A New Search Engine

- ## Hyperfocus on this issue:

- NOT a "new Google" - but instead keeping the focus targeted on local shopping
- Want to developing a localized transactional search engine

- WHY? Finite and exists

1) Limits the "endless" aspect of the internet in terms of information or actual websites; businesses are finite with some barrier to entry (limited and known geographic area, need to have the product)

2) They do exist

Google SEO: The Bias Against Small Biz

[3] Looking at Google: A BIG Budget Bias

- ## Google Bias: Quality is subjective

- Google claims they strive to give you the "highest quality" results - but I found an article from an independent search engine company that talks about the inherent bias in Google's search

- Google is "playing god" by defining high quality - which then defaults to **advertising and bias**

- ## Two Examples: Search engine bias and advertising bias

- Google knows I'm a football fan, not an oceanographer, so they give me the **Miami Dolphins instead of the animal**
  - Hints at a bigger issue of a filter bubble
- Another example is **WebMD vs National Institutes of Health**
  - Advertising driven sites like WebMD often outrank government sites like National Institutes of Health. This is because WebMD is better at getting people to read through their content and click around.

[4] Defining "Relevance"

- ## Google: Relevance

Google says they look at several factors to decide what is valuable:
- Location, Search history, and Search settings are important, content loads speed

- ## So What → They are prioritzes factors that aren't about having the product or quality for a customer

I don't care if Nike's website runs the fastest, I want to run the fastest in my track spikes!
- I want to know the QUALITY or VARIETY of track spikes, NOT if there site loads quickly!

[4.1, 4.2] HBO Max Google Shopping Ad

- ## Google Shopping Ad - Fake & Google gets a cut

They are likely created for marketing purposes only and were not readily findable after cursory searches.

Google gets a cut

[4.3] ChatGPT

- ## READ SLIDE ALOUD

Even when Bing was pressed on the issue, it ended the conversation

[5] SEO as current metric

- # This leaves local businesses looking to Search Engine Optimization

- Google's definition is hurting the consumer as its hiding quality small businesses that don't want to play the "SEO" game
  - Official Google Documentation says to hire an SEO expert
  - Small businesses can't do this since they are busy RUNNING THEIR SMALL BUSINESS

This shouldn't be the solution

[6] SERP Assignment

- # Empiricial Data - Running Stores in Philly

We see this in action with different Search Engine Results Pages (SERPs) (Miami Dolphins vs Dolphins) but also across search engines

Search Engine Theory:

[7] Theory:

- # [EXPLAIN THE CHART]
  - "A transactional search query is a query that indicates an intent to complete a transaction, such as making a purchase"

Three types within traditional search engines philosophy - I developed a sub caterogiy within one of them

  - Many **local searches** are transactional as well → "Dever wine shop"

- # Also different formats of search:

Key word to give 10 Blue Links vs Topic hierarchy

[8] Topical Tangent:

- # Informational Searches rapidly changing

From "10 blue links" to actual answers → why ChatGPT is so good
  - Google had already actually been attempting this with their "Knowledge Panels"
    - Google has been on it: *Google is answering the question directly*

  - GOOGLE: **Knowledge graph** built, used Wikipedia and others to augment their indexing (find answers to questions) → giving panels that are Wikipedia like (information retreival made more direct)
  - vs Large **Language Model, NOT knowledge-based, more like fancy autocomplete**
    - ChatGPT being factual is a side effect for the model being trained on so much data
      - (not designed to give a built a factual answer)

- ○ Chat-based, interactive component - makes it seem like advanced Google
- ○ *ChatGPT not task trained, can do a lot of things well → emergent knowledge that comes out of nowhere ("generalized learner")*

[9] Search Engine History

Yellow Pages

- If you had a business, you just had to list yourself in the yellow pages, it was the single and most basic and important way to advertise yourself. If you missed out in being printed in the annual yellow page directory, it was as if you didn't exist for the entire year.

YAHOO

- QUOTE: Yahoo implemented hierarchical browsing "improves precision by first defining and narrowing the domain for searching" It does so by collocating (inclusion) and partitioning (exclusion)
- THREAT:
  - ○ Fails when there's a massive amount of possibilities
  - ○ QUOTE: The essential feature that led to the explosive growth of the web – decentralized content publishing with essentially no central control of authorship – turned out to be the biggest challenge for web search engines in their quest to index and retrieve this content

Google

- Page Range – Web like a social network, each page has friends → Link Analysis (Link in, Link out

THREAT:

- Web Assumptions:
  - ○ 1) The anchor text pointing to page B is a good description of page B
  - ○ 2) The hyperlink from A to B represents an endorsement of page B, by the creator of page A.

MetaSearch Beta: NearBuy

[10] Goal:

- ## GOAL: [READ SLIDE]
- ❖ *Build a (prototype of) a localized transactional search engine product that is SEO-proof, mirrors the trust and fairness that characerized the printed Yellow Pages, ensures inclusivity of all local businesses, and delivers a universal SERP tailored to specific geographical areas.*

WHY?
- *Mission Statement:*

- ## By developing a more comprehensive and unbiased search directory, we can help people to more easily find the products they need and make informed decisions, free from the constraints of traditional search algorithms that have rapidly become commercially "gamed", full of biased, and questionable inclusivity.

[11] ISSUE:

- ## This is difficult there's not a centralized database of stores and their goods
  - Business Permit Search/Business databases → not comprehensive or updated regulated
  - Could scrape Google Maps but that doesn't classify them nor list the types of products/brands they have
  - IE:
    - Working from the "store down to→ product" direction doesn't tell you what DICKS sells
    - Also, I didn't want to just rerank Google

[11.1] SOLUTION:

- For prototype - found interesting work around to develop off of → Store locators!

some brands have their own product/brand locators

[11.2] PBR Example

- Example: PBR [EXPLAIN]

- Assumption: Accurate and Comprehensive
  - **Assumption:** This prototype model assumes the store locators are **accurate and comprehensive** (something we're just going to have to live with for the extend of this independent study)

[12] Prototype Info

- Workable: brands as surrogate for product

Still not product specific (Assumes if you have Asics, you have their track spikes, not their running shoes)

[12.1, 12.2] Vision

Mascot: Oliver Towne
- Embodies the values:
  - Oliver Town - Product "All over town"
  - Penguin - flightless bird, must go somewhere local
  - Yellow Coloring - paying homage to Yellow Pages
  - Black and white - as simple as back and white

[12.3] Live Demo

SHOW WITH SAMPLE ZIP CODE

[12.4] Semester Learnings **(Preface)**

TECH SIDE: HOW IT'S DONE
- Inspecting for certain elements via HTML → using selenium to select and python to loop

● **Single Page Application**
  - Update the page without the URL changing

[12.5] Semester Learnings

WALK THROUGH CODE

[12.6] Semester Learnings

SeatGeek - Explain
- FuzzyWuzzy was actually developed by seatgeek to match event tickets from different corners of the internet - based on Levenshtein distance
  - They found identifying whether two ticket listings are for the same real-life event can be a consistently frustrating issue.
  - The library provides various fuzzy string matching routines, including partial string similarity and out of order string construction matching, to programmatically compare and match event listings accurately

[12.7] Semester Learnings

Explain FuzzyWuzzy Code Example

- The intuition here is that because the SORTED_INTERSECTION component is always exactly the same, the scores increase when (a) that makes up a larger percentage of the full string, and (b) the string remainders are more similar. In our example

MUCH LIKE FINDING A LOCAL STORE AROUND THE CORNER → I was able to find something really valuable from something a lot closer to me than I realized

[12.8] Recap

- Built modern day yellow pages

===

- Reflecting on this experience, bigger idea came to mind

Reimagining Product Search Entirely

[20] Idea

- ● Still didn't actually find the track spikes

All this lead me to a bigger idea: still how to I find who physically has track spikes in the physical region??

[20.1] The BIG idea

- ● [READ SLIDE]
  - ● *Similar to Amazon's "Other Sellers on Amazon" feature but more comprehensive that directly taps into real-time inventory of all retailers.*

[21] Imagine a Side

- ● Imagine a Site → Walk through graphics

Extension of NearBuy - but even more, Comprehensive, Centralized

[22] Ideal System

- ● [EXPLAIN SLIDE FLOW]
- ● Going beyond listing your business in Yellow Pages → actually listing what people care about, the PRODUCT

- ● Key it off of Universal UPC

[22.1] Corporate Sponsor

- UPC's do exist → system is in place to implement this realistically

- How to implement:
  - *Microsoft/Shoptify/Visa/… gives out a free POS inventory system that connects to a large cloud base (retailer CAN NOT opt-out) that makes auction style portal to show all available (in-stock) locations across the US of that exact color and size item*

- Shopify already has everyone's individual data

[23] Visionary:

- **EXPAND UPC system to services too**
  - Recent: Scooter repair, hardware store
- Becomes a MAP for buying anything and everything near by
  - *This map tells you if you live in a 15-minute city*

- Bottom-up solution for products
  - Distributed social
  - Influencer, data-generation

People who are selling products (PBR), right now, they aren't being discovered on Google

[24] Showcase

- Closest is specilized bikes website
- Find-in store

- Find stores

[READ SLIDE]

- This is what small businesses are resorting to