Phase Five (Capstone)
Capstone Development Proposal

Your Name: David Schenker

Capstone Project Name: **Spotify Playlist Curator**

Instructions:

1. Make a copy of this document: **File -> Make a copy**.
2. Fill out all relevant information.
3. Change sharing permissions to **"Anyone with the link can view"**.

NOTE: This proposal template version is unique to **data science** students.

📍 PROJECT IDEA SYNOPSIS 📍

Why are you building this project? What's the motivation behind it? Identify and clarify the domain interest that this project focuses on, as well as whether or not the project is intended to solve a specific problem, address a particular need in the world, or emulate a preexisting company/tool.

Write your summary statement in the following section.

This project aims to build a playlist curation system for Spotify users by analyzing song attributes and metrics to find similar and comparable songs that the user would enjoy. The goal is to enable users to find new music and quickly and easily create fresh playlists.

Your summary statement is 1-2 sentences that objectively describe the function, intention, and direction of your project – what your expected end result is supposed to look like and/or accomplish.

Remember: you are still operating in the realm of business-facing data science, and as such, your work *needs* to be applicable to the realm of business value. Your summary statement comprises a summary of why a specific business domain would find your intended work interesting and potentially hireable. In short, this statement should serve to justify the distinction between this project being an industrial-grade exploration vs. a non-pragmatic technical tangent.

Your summary statement should be short, simple, and effective for contextualizing the next three weeks of your project development process.

Which of the three possible specialties will you choose for your capstone project: an insights intensive, a research intensive, or a product intensive?

In other words, are you choosing to leverage analytical and inferential insights from industrial data for immediate interpretability across a business? Or are you choosing to dive deeper and explore the applicability of a bleeding-edge field of data science within a domain-of-interest for the purposes of discovery? Or are you choosing to focus on domain-facing insights to construct software with an immediate application in industry for the purposes of invention?

Choose your project direction-of-specialty at this time.

- ☑ ~~Business Insights Intensive~~
- ☑ ~~Research & Development Intensive~~
- ☐ **Software Product Intensive**

If you are choosing to specialize in the **Business Insights Intensive**, then your goal will be to carefully curate data representing real-world information and construct a pipeline that deeply explores, investigates, describes, and infers patterns while impressively communicating business insights to stakeholders and general audiences.

If you are choosing to specialize in the **Research & Development Intensive**, then your goal will be to produce a sophisticated in-depth exploration of custom algorithms and predictive technologies at the fringes of data science and demonstrate potential future opportunities in your domain-of-interest.

If you are choosing to specialize in the **Software Product Intensive**, then your goal will be to produce an end-to-end data science pipeline that can be immediately applied into a prototypic software product allowing for interactivity with prior and new data.

What does the minimum viable version or V1 (a.k.a the "skateboard") of your product idea look and feel like? What about the V2 project (a.k.a. the "bike")? What about the V2 project (a.k.a the "car")? What about future versions (a.k.a. the "truck", "rocket ship", or "space station")?

In the following section, identify and describe relevant details for your project's rapid prototype state (V1), its first stable development version (V2), and its second stable development version (V3).

V1: Create a Model that can take in a single song and be able to return that song's attributes and metrics

V2: Create a Model that can take in a single song and be able to return a single song that has similar song attributes and metrics.

V3: Create a Model that can take in any number of songs and be able to return any number of songs that have similar song attributes and metrics

Your MVP should be designed to be completed in 1-3 days of work and should encompass a rapid traversal of as much of the data science design process as possible.

Your V1 project should further explore the data science design process but begin to hone in on your chosen areas of focus as it pertains to either a research or product intensive direction.

Your V2 project should be an evolution of your V1 development into a stable foundation either impressive in its cohesive structure, depth of research, or usability as a product.

What known technologies are you planning to use to build the initial vision of your product? What new technologies are you planning to use to fully actualize your product vision? At what point is it necessary to integrate new technologies into your stack?

List key technologies needed to develop your product.

Python, Pandas, TensorFlow, Scikit-Learn, Spotify API

You don't need to jot down every minor piece of your technical stack, but it's important to identify major languages, dependencies, frameworks, libraries, utilities, and services needed for your product to function… even if you already know it. This will help as you design new versions of your product to better diagnose bugs and incompatibilities before they happen.

For one-week sprints, detailing roughly expectant deliverables on a day-to-day basis is critical to your success. While these scopes don't have to be (and oftentimes will never be) very accurate, you will notice that they will certainly become more precise over time as you become better with scoping more achievable and comfortable challenges and foci with each passing day.

Clearly approximate and identify daily deliverables to better scope your Week One objectives.

- ☐ Monday, September 16:
- ☐ Tuesday, September 17: Project Proposal + EDA with Spotify API
- ☐ Wednesday, September 18: EDA with Spotify API + Creating the Base Model
- ☐ Thursday, September 19: Base Model Tuning
- ☐ Friday, September 20: Completed V1

These deliverable statements aren't meant to be either accurate or detailed; rather, they'll serve as benchmarks to help you identify how slowly or quickly your product development process is progressing and if rescoping, testing, backtracking, or soft/hard pivoting is necessary.

[Required for insights-driven intensive projects.]
Write up a business analysis proposal for initial explorations and investigations into your data source.

In this capstone phase, a "business analysis proposal" is effectively an extension of the proposals you did in early phases when scoping out an Exploratory Data Analysis, except now it must include more nuance and detail as your data sources, heuristics, and applications may be more ambiguous.

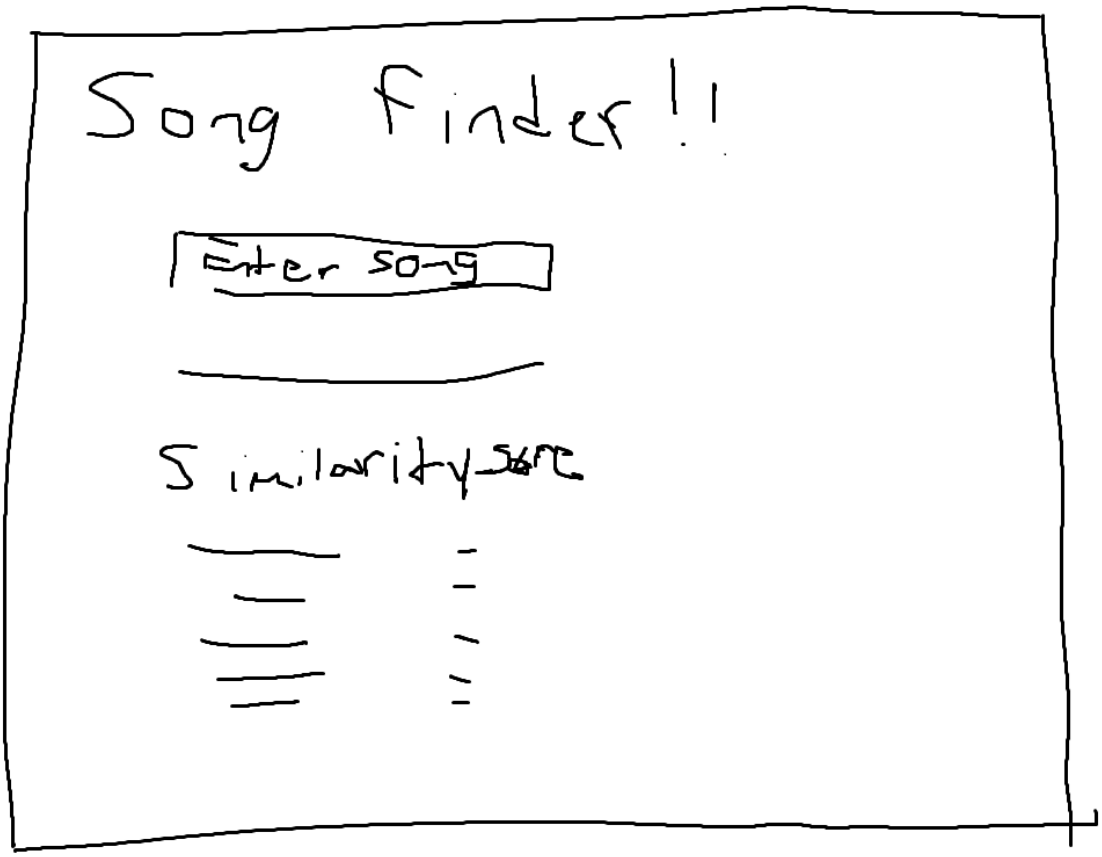
As such, your proposal should comprise answers to the following questions:

- **Stakeholders:** who are the people affected by this project? What are their needs and expectations?
- **Metrics:** What measurements will be used to measure the data? What key performance indicators are necessary to use?
- **Data Sources:** What data sources will be used? Where is the original data located? How will you access the data?
- **Existing Solutions:** What preexisting solutions or tools are there that may partially/fully address the problem you're trying to solve?
- **Resources:** What external resources may be useful to research and better understand the data's domain?
- **Describing Patterns:** What sample questions can serve as a foundation for performing descriptive analyses across the target data?
- **Inferring Trends:** What sample questions can serve as a foundation for performing inferential analyses across the target data? What hypothesis tests may prove useful for defending/evaluating these assertions?
- **Ethics:** What ethical implications exist for investigating our business data?

[Required for research-driven intensive projects.]
Mockup and produce model card(s) for your intended predictive algorithm(s) that you intend to research in this project.

For research and development of new algorithms - or preexisting algorithms for novel and unique purposes - the inclusion of model cards is extraordinarily important to communicate useful information such as intended usage, limitations of training/testing, specific sources for data, columnal metadata, imputation/cleaning techniques used, etc.

[Required for product-driven intensive projects.]
If applicable, upload rough diagrammatic wireframes and/or software architectural diagrams to represent your product's layout and schema.



Wireframes don't have to be immensely detailed works of art – they serve as visual models for referencing what you want the application to look like and how you want users to navigate it.

Software architectural diagrams don't have to be incredibly sophisticated blueprints of your final product either – they serve as visual models for referencing how you want the overall application or service to be broken down and how individual pieces and components interface with one another to give rise to the service as a whole.

Neither wireframes nor software architectural diagrams are set in stone upon creation, and can (and should) change as you develop your project further; they are meant to provide solid foundations for inspiring development in the right direction.