groov.

Data Driven Medicinal Cannabis

Sai Anantapantula, SSE | WH (FNCE), sanantap@wharton
Cole Fairchild, SSE | WH (OIDD), cfair@seas
Will Jansen, CIS | WH (FNCE), wjansen@wharton
Fatima Koroma, SSE ,koroma@seas
Antaures Jackson II, CIS | WH (FNCE), ajj4311@wharton

Advisors and Mentors:
Siddharth Deliwala: deliwala@seas.upenn.edu
Jan Van der Spiegel: jan@seas.upenn.edu
Donovan Schafer: doschaef@sas
Franklin Caldera: Franklin.Caldera@pennmedicine

Table of Contents

II. Executive Summary
III. Overview of the Project
IV. Technical Description
V. Self-learning
VI. Ethical and Professional Responsibilities
VII. Meetings
VIII. Project Schedule
IX. Discussion of Teamwork
X. Budget and Justification
XI. Standards and Compliance
XII. Discussion and Conclusion
XIII. Business Analysis
  a. Executive summary
  b. Value Proposition
  c. Stakeholders
  d. Market Research
  e. Competition
  f. IP If Any
  g. Cost
  h. Revenue Model
XIV. Appendices

II. Executive Summary:

One of the primary issues with the current medical cannabis industry is how opaque the product selection process can be. Customers don't have awareness around how cannabis varieties impact them pharmacologically, and Retailers don't have the education or tools to route patients to informed treatment at scale. The state of Pennsylvania requires a pharmacist to be on staff at a retailer for this nuanced guidance, but availability and knowledge of this resource are limited. Groov is stepping in to solve this issue by providing an experience logging and education tool for the patient, creating a data solution for the dispensary to inform product guidance (as most doctors that prescribe medical marijuana do not recommend a specific chemical compound). The platform we've created consists of two components: (1) Journaling app for medicinal cannabis users to track their experience with each product's chemical makeup, with insight crowdsourced across the entire user group (2) An AI mechanism that uses this crowdsourced data to augment the retailer's recommendation behind the counter with profiled information on a user's desired chemicals and effects.

Noting a platform challenge in creating data inflow, we've designed a scheme to test boilerplate learning models on synthetic datasets we've generated based on limited survey input and current academic research gathered surrounding various ailments where certain chemical compounds have been shown as proven methods of relief. Our intention is to illustrate how the platform can facilitate insights at the patient and retailer level informed by synthetic taste profiles. On-platform, the patient will be able to journal a cannabis experience with a product-specific chemical mix through rating the impact of key relief factors. Providing easy routing to sources of truth on cannabis science and wellness is core to the value that will drive user interaction. ChatGPT is a conversational intelligence tool for direct Q&A, alongside content from Greenbridge Health to explain cannabinoids, terpenes, and other scientific cannabis concepts.

In addition, the recommendation system will be targeted towards retail dispensary owners as they are the primary point of sales for medical marijuana products and they do not have strong information/knowledge related to the correlation of certain strains' chemical composition and their effects for treating a variety of illnesses. Based on some of the data and anecdotal experiences (from the surveys we sent out), there seems to be a correlation between chemical strain composition and the effectiveness/efficiency of treatment for certain diseases as well as interactions e.g. some terpenes create difficulties for people with migraines or they are more effective at treating inflammation. As of now, we have created a front-end to allow users to journal their usages, share their empirical ideas (i.e. what works for them) connected to a database backend. Our next steps for the project include building out a front end for the retailers to provide recommendations given a patient's background as well as a tool for helping users to better calibrate what they are using.

III. Non-Technical Overview of the Project:

We aim to tackle the challenge of mapping chemical strains and compositions of marijuana by implementing a solution on the retail level. By quantifying and gathering data through crowdsourcing, we can improve the quality and personalization of strain recommendations for patients. The more users that use this platform, the better the recommendations will be, a positive feedback loop. We will monetize this data by selling the data insights to retailers to augment their own individualized recommendations as well as selling anonymized data to a variety of participants in the market ranging from growers to doctors as it could help them improve parts of their business. For example, trends in strains for treating certain diseases can help a grower decide what to plant or what a retailer should put on their shelves for their target markets. As marijuana legalization and acceptance as a medicine increases, now is the perfect time to develop a data-driven recommendation system to capitalize on the expanding market.

IV. Technical Description:

Learning Models

The desired prediction of our learning models are to extract greater insight about what drives certain relief responses in patients using medicinal cannabis and to be able to recommend chemical profiles that suit specific patients' needs.

Our data categorization is as follows:

Demographic Information
- Age
- Gender
- Ailment for which patient is prescribed medicinal marijuana

Preferences for relief
- Desired Pain Reduction
- Desired Energy Level
- Desired Mental Clarity
- Desired Psychoactivity

Single Use Consumption
- THC% Consumed
- CBD% Consumed
- CBN% Consumed
- Terpenes Consumed
- Mode of Consumption

Single Use Impacts
- Pain Reduction
- Energy Level
- Mental Clarity

- Psychoactivity
- Overall Effectiveness in Addressing Ailment
- Side Effects Present

Prediction Values
- Optimal THC%
- Optimal CBD%
- Optimal CBN%
- Optimal Terpene Inclusions

Given that the impacts of medicinal marijuana can be different for different demographics and across ailments, it is primarily important to identify meaningful clusters of similar medicinal cannabis users within the dataset. This can be achieved through a K-Means Clustering unsupervised learning algorithm whereby we compute clusters by randomly initializing k cluster centroids of significant distance (K-means++ initialization), and iteratively assign points to clusters by their smallest euclidean distance of features that are composed of relief preferences and demographic information from the above data categorizations and re-estimating cluster centroids based on the data within each cluster until convergence. We are then optimizing the objective function:

$$argmin_S = \sum_{k=1}^{K} \sum_{z \in s_k} ||x - \mu_K||_2^2$$

This provides us with meaningful clusters and a base level recommendation by taking the centroid average THC, CBD, CBN, and upper quartile terpene inclusions for new users by assigning them to the cluster with shortest distance to centroid. It additionally provides a basis for data split to run k, linear regression machine learning models that are specific to each cluster.

Within each of these clusters we can run a linear regression machine learning model with the matrix X representing the feature set composed of demographic data and single use consumption data from above and the matrix Y representing the corresponding single use impacts from above. We are then optimizing the loss function for Beta:

$$L(\beta;\, Z) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \beta^T x_i)^2 = \frac{1}{n}||Y - X\beta||_2^2$$

And so we can compute the gradient:

$$\nabla_\beta L(\beta;\, Z) = \nabla_\beta \frac{1}{n}||Y - X\beta||_2^2 = -\frac{2}{n}X^T Y + \frac{2}{n}X^T X\beta$$

Since the closed form solution is computationally expensive, we can use gradient descent with learning rate α, continually recalculating the gradient until $||\beta_t - \beta_{t+1}||_2 \le \epsilon$ with ϵ ~ 0.0001:

$$\beta_{t+1} \leftarrow \beta_t - \alpha \nabla_{\beta} L(\beta;\ Z)$$

Data Synthesis

Given difficulties in data collection in training the model, and limited responses to survey data that we have put out, we have had difficulties in getting the necessary data on which to run our models. This is a classic issue in network growth, whereby the validity of the model depends on users, but users are difficult to acquire without a well functioning model. For this reason we have employed some randomization and data synthesis to extrapolate limited response data around some Gaussian distribution noise.

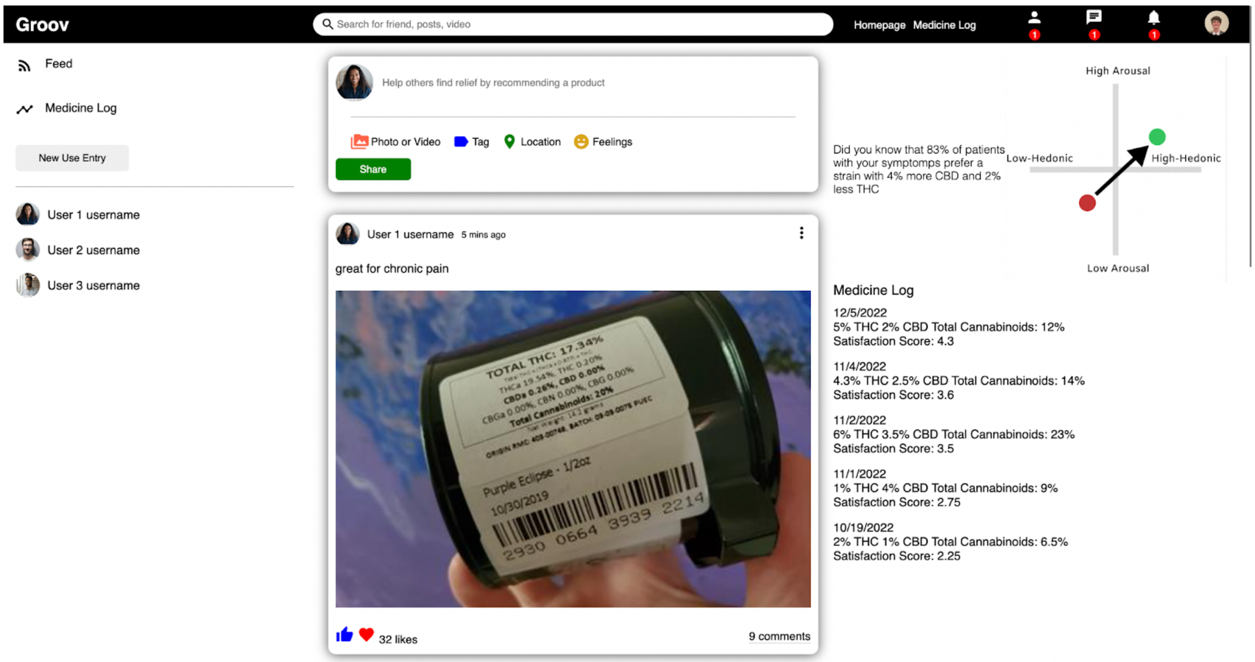
The algorithm we are considering to use for this data simulation is called Variational Autoencoder (VAE).

1. The encoder itself maps an input x to an unobservable variable z essentially the algorithm is mapping from our highly dimensional input data to a lower dimensional space
    a. We can think of the probability distribution for z as follows: $Q(z|x) = N(\mu(x),\ \sigma(x))$
    b. Q is the encoder probability density distribution which is typically represented by a normal distribution and then the parameters of the normal distribution include the inferred mean and standard deviation
2. Once a data point x has been mapped to the lower dimensional space at the point z, a random sample is created from the distribution $Q(z|x) = N(\mu(x),\ \sigma(x))$
3. The decoder part of the algorithm maps the lower dimensional space variables z to generate new data points since this is mapped and then re-mapped, we refer to the new data points as x' i.e. $x' = P(x|z) = N(\mu(z),\ \sigma(z))$ - essentially the posterior of the encoder probability distribution function
4. The model for VAE is trained similar to most of the other machine learning models in terms of minimizing a loss function i.e. the difference between x and x'. The loss function is the reconstruction loss (how well the data generated matches to the input data point) and then also the regularization loss (i.e. how much information we lose from dimensionality reduction)
    a. $L(x, x') = -E[\ln P(x|z)] + KL(Q(z|x)\ ||\ P(z))$ where $P(x|z)$ is the decoder distribution, KL is the Kullback-Leibler Divergence, $Q(z|x)$ is the encoder distribution, and $P(z)$ is the prior distribution over the lower dimensional space
5. The model like most Machine Learning models are by optimizing to reduce the loss function via gradient descent
6. In order to generate simulated data, we just pass the sample from the distribution for the lower dimensional variable through the decoder network to generate our data points.

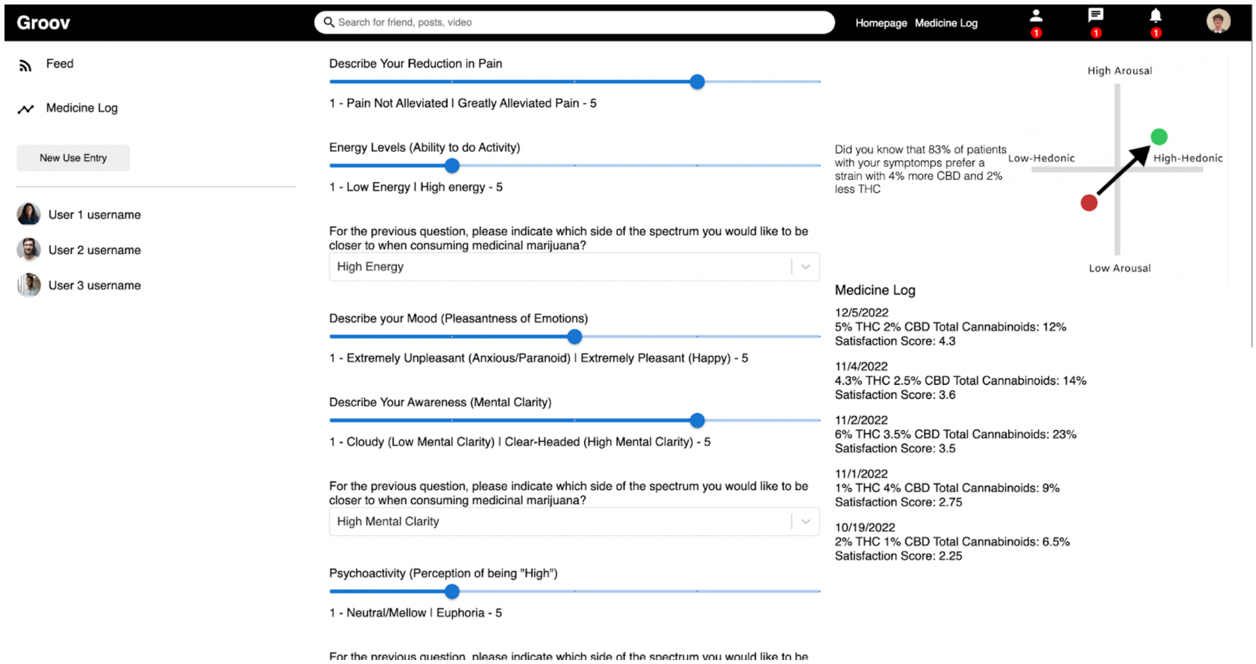
Web Architecture

The web architecture for our platform was built using React for the client side front-end. The backend of our application is run through a server built in Node.js and Express with the backend database hosted on a MongoDB cluster with an AWS instance.

User Front End



The user end platform is displayed above and was created using the React framework in javascript. The main libraries used for the creation of the front-end application were from @mui.icions-material and routing through react-router-dom from npm.js. The platform encapsulates several key functionalities that together create a functional interface for medicinal cannabis users to iteratively improve their treatment over time. The application allows medicinal cannabis users to connect with one another as with other social media platforms to share anecdotal relief stories and advice for others who suffer from similar ailments and utilize medicinal cannabis as a remedy. The platform additionally provides links to resources for medicinal cannabis education, and live consultation services by medicinal cannabis professionals. The application allows users to log their medicinal cannabis use over time by responding to a series of questions as seen in the following image.



This data collection provides value to the medicinal cannabis user, as they are able to formally track their consumption over time to determine histories of most effective outcomes. This data collection is a crucial aspect of our project, as the user provides us with answers to questions regarding pain reduction, energy levels, mood, mental awareness / clarity, psychoactivity, overall efficacy, side effects, mode of consumption, and major chemical profile characteristics such as THC, CBD, CBN, as well as common terpenes in medicinal cannabis products. These data points will provide the inputs to our machine learning models that drive a recommendation algorithm for new users to determine which products will work best for them individually.

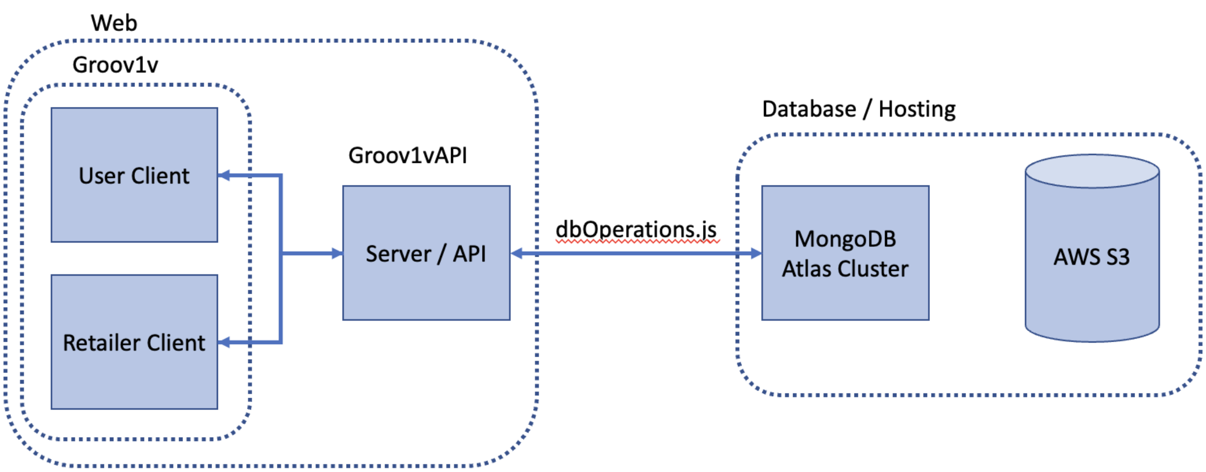
Retailer Front End

The retailer front end provides the other aspect of our platform, enabling in-store medicinal cannabis retailers to provide more tailored recommendations to their customers rooted in the models that we train on the user end data. The retailer platform, similar to the onboarding for the user-end flow, asks a few short demographic, ailment, and desired medicinal impact questions that a customer could answer in just a few minutes in store that then utilizes our model to yield a product recommendation. This recommendation provides a desired chemical profile for major cannabinoids including THC, CBD, CBN, terpenes, and total cannabinoid levels. The retailer user flow is such that they

1. Login
2. Provide answers to demographic and desired impact for relief profile
3. The retailer submits the responses and the platform yields a recommendation profile

Back End

Data collected from this application communicates via a REST (Representational State Transfer) API we have developed specifically for this application use case. This REST API communicates with a REST Server that concurrently runs with the front-end application to handle HTTP requests over the internet. This server communication then facilitates data flow to a database hosted with MongoDB on AWS via a set of database operations that we have coded specifically for this application implementation. This allows for two-way dataflow and communication between the local web application and our database hosted on the backend, with proper permissions to ensure anonymity for requests from the retailer portal and standard permissions for requests from the user portal.



V. Self-learning:

There were several key areas of self-learning and technical development that necessitated gaining knowledge beyond the foundations of the curriculum at Penn. Building upon knowledge gained in classes like CIS 350, our project required the development of a full-stack web application built with React, Express, Node.js, and hosted in MongoDB with database functionality. CIS 350 provided the bedrock from which we built basic knowledge of building this sort of application, but our specific application required extensive research into documentation of various JavaScript libraries to build out our front-end UI, and frameworks for safe transmission of HTTP requests across the internet to send data to and from our database. With regards to the machine learning algorithms leveraged in our recommendation model, there was a greater degree of self learning beyond the classroom. Building upon the foundations of machine learning applications from classes like CIS 4190, we had to expand our knowledge and application of base models for our use case. The variational autoencoder framework for data synthesis builds upon many of the basic neural network architectures taught in class while also leveraging statistical knowledge and linear algebra manipulation from classes like MATH 3120 as well as STAT 430 and STAT 431. The VAE procedure required fine tuning of the number of dense neural network layers best suited for our data. The dimensionality reduction and re-mapping of lower space data to our original feature space was an interesting area of broadening the scope of our linear algebra and machine learning knowledge in the context of reconstruction loss relative to typical loss functions. Additionally, the loss functions of our linear regressions among the "quasi-labeled" data that we had, and the choice of which parameters to include in our Beta matrix to minimize the measured loss was an area of exploration for us in our technical learning. We had to augment the basic frameworks of K-means clustering and KNN classification learned in the classroom to our more complex dataset given the number of different data points we could consider in formulating the models that we did. The most important feedback that we received this semester in regards to these learnings actually came from advisors we sought in the emotion research space. Choice of data responses with which to build our models around provided a lot of ambiguity in how we wanted to categorize data that we would receive from users through the platform and through surveys, and drilling down to the questions set in our final product was an iterative and confusing process that required a lot of changes of direction. For something like the efficacy of medicinal marijuana, there is not one dimension of outcome that fits every patient, which is partly what makes this project so interesting. It is unlike many other medications that seek to address a specific need but rather a broader scope more suited to unsupervised learning where we can better understand which strains are better for pain reduction compared to anxiety compared to appetite compared to sleep outcomes and so forth. This is a highly complex interdisciplinary problem that requires far more work beyond the scope of our senior design project, though we have been energized by the small portion we were able to engage with these past two semesters.

VI. Ethical and Professional Responsibilities:

On a societal level, medical marijuana has become increasingly popular as a means of treating a variety of ailments ranging from insomnia to anxiety. With the rising popularity of the drug, there has been ample discussion both in favor and against the widespread adoption of medical marijuana as a treatment option. Socially, there is worry that the popularization of the drug medically will directly correlate to increase usage of the drug recreationally. In fact, some states, like Missouri, progressed from the legalization of the drug medically to the legalization of the drug recreationally. Economically, medical marijuana has several potential benefits. After studying trends in Colorado and Washington since the legalization of medical marijuana, it has been shown to result in buoyant tax revenues and millions in tax revenue as of 2021. An increase in medical marijuana dispensaries and nurseries also creates thousands of jobs for Americans while also opening up a variety of investment opportunities.

The societal controversy around medical marijuana creates ethical concerns for our product as Groov promotes the usage of medical marijuana. With increasing THC content in medical marijuana, the potential long term effects of the drug is unknown. Based on prior history, it is unlikely that severe long term effects will be observed, but it is crucial to note that the app would promote the usage of the drug despite not having complete knowledge of the long term effects of the drug. While Groov would take into account the individual's unique medical history, every patient has a genetic makeup and biochemistry that reacts to cannabis differently, which can lead to the recommendation of a treatment that is not effective or even harmful. Additionally, these tools may subvert the role of medical professionals by allowing patients to self-diagnose and self-medicate without proper oversight or guidance. An ethical concern that is crucial for us to address is the potential for misuse of the app whether it be through recreational use or being used by underage consumers.

A professional responsibility would be to partner with retailers and medical professionals to educate the user on substance abuse and potentially create oversight or overuse warnings. Additionally, it is necessary to create a measure that would ensure that the application can only be used for medical marijuana usage and not for recreational usage. This concern can be addressed through the user being forced to scan their medical marijuana card, as well as a clear form of identification. Including proper identification in the intake process would also allow us to ensure that all users are of age, and are using the product responsibly.

VII. Meetings:

During the first semester, we spent a lot of time trying to understand the product market as marijuana, and its potential interactions with an individual's endocannabinoid system, is complicated and largely misunderstood. As a result, we met with several subject matter experts like dispensary pharmacists and full service medical cannabis facilities. These informational interviews would take place around twice a month. As a team, we met once every other week to discuss progress and team deadlines. We also met with our initial advisors, Donovan Schafer and Franklin Caldera, twice each over the course of the first semester. During the second semester, majority of our meeting time was amongst the team as well as with our new advisors Jan Van der Spiegel and Siddharth Deliwala. The team meets an average of once every week, with increased meeting time around project deadlines. We also increased meeting time with our advisors by meeting once every two weeks to discuss our team's progress and try to keep us moving forward. While the time spent meeting with subject matter experts decreased, they will pick back up as we near the end of the semester in hope of synthesizing data to test our project.

VIII. Reflection on the project schedule and level of achievement:

	Fatima	Cole	Sai	Will	Antaures
Finalize server back end development and deployment (ensure endpoints with correct HTTP request handling)				1/31/23	1/16/23

Find Qualitative Correlations in User Intake Survey			1/16/23		
Finalize Intake Questions based on the User Survey	1/21/23	1/21/23	1/21/23		
Set Up Retail Front End Development Environment & Initial Mockup					1/28/23
Refine Front End Development to incorporate OCR and intake failure pipeline				1/31/23	1/30/23
Integrate outstanding items related to OCR and intake failure pipeline with Retail Front End					2/12/23
Finalize intake variables and generate dummy data set	2/27/23 Intake Vars	2/27/23 Intake Vars	3/12/23 Data Gen		
Connect dataset to MongoDB					2/3/23
Finalize Retailer Front-End Design (Figma)	2/24/23	2/24/23			2/24/23

Finish Front End Devs Integrate User Permissions				2/24/23	
Finalize Integration of Retail Front End w/ Shared Platform Backend through MongoDB					3/1/23
Ensure Completion of any outstanding Development Items Prior to Testing				3/3/23	3/3/23
Test Application with UI, End-to-End, and Integration Testing				3/3/23-3/10/23 (Flexible based on the status task above)	3/3/23-3/10/23 (Flexible based on the status of task above)
Finalize Application and Design based on testing	3/24/23	3/24/23	3/24/23	3/24/23	3/24/23

IX. Discussion of teamwork:

The team has been divided between three groups, a data team, a design team, and a software development team. The data and design team consists of Fatima Koroma and Sai Anantapantula while Will Jansen and Antaures Jackson work on software development. Cole Fairchild and Fatima also make up the design team and are responsible for the UX design of the final application. Cole began by researching subject matter experts for the team to meet with in order for every member to gain a thorough understanding of the project. Sai took charge of the data team as he has the most extensive experience in the area. Getting the wheel moving, Fatima created a user intake survey and posted it on subreddits to collect data over the course of winter break, and Sai has begun analyzing the collected data to see if there is a correlation in the way patients with similar ailments described and rated their experience. Will finished the initial draft of the front-end development to be developed and refined by him over the course of the second semester. During the second semester, Sai refined the user intake questions to better correspond to our final project, worked on data synthesization, as well as analyzing the data we were able to collect. Fatima assisted him by posting the survey on Philadelphia Medical Marijuana reddit boards. For the software development team, Will led the software development of the product by finalizing the entire front end as well as planning out the backend configuration. The whole team worked together to discuss and refine our project idea to include multiple dashboards.

X. Budget and justification:

Our initial budget was primarily focused on hosting the backend using MongoDB and covering associated cloud service fees. However, since we did not deploy the platform during the project timeline, these expected expenses were not incurred. As a result, the budget allocated for these purposes were not drawn upon.

XI. Standards and compliance:

There are several industry standards associated with the development of our project. These standards primarily stem from the medical cannabis, data privacy, and software development sectors. Adhering to these standards is essential to ensure compliance, maintain user trust, and deliver a high-quality product.

As our platform operates within the medical cannabis industry, we have made an effort to comply with the regulatory frameworks established by various jurisdictions particularly in Pennsylvania. These regulations include guidelines on patient registration, product recommendations, data collection, and data sharing. Staying up-to-date with these regulations and adapting the platform to different jurisdictions is crucial to maintaining compliance and avoiding legal issues.

In providing value to patients and dispensaries, our platform is positioned handle sensitive user information, such as medical conditions, cannabis usage patterns, and personal preferences. As a result, it must adhere to data privacy and security standards like the Health Insurance Portability and Accountability Act (HIPAA) in the US. This standard dictate how personal data is collected, stored, and shared to protect user privacy. Groov must ensure robust security measures are in place to safeguard user data and comply with these regulations. As we have synthesized the vast majority of the data we worked with during the semester this is not something we strictly adhered to, but

Additionally, we followed industry best practices in UX design and accessibility to create and inclusive and user-friendly platform. The Web Content Accessibility Guidelines (WCAG) provide a comprehensive framework to ensure that digital content is accessible to people with disabilities. By adhering to these guidelines, we have a more inclusive platform that caters to a more diverse user base.

XII. Discussion and Conclusion:

Throughout the course of our project, we have made significant progress in developing Groov into a comprehensive platform designed to streamline the medical cannabis selection process for patients, budtenders, and retailers. Our initial focus was on understanding the medical cannabis market, its challenges, and the unique requirements of its users. Through thorough research and regular consultations with subject experts, we refined our solution to cater to the specific needs of budtenders and patients. This solution is in the form of a platform that consists of two main components (1) a web-application that allows patients to track and share their experiences with different cannabis products, and (2) a recommendation system that utilizes the gathered data to help retailers provide personalized product guidance.

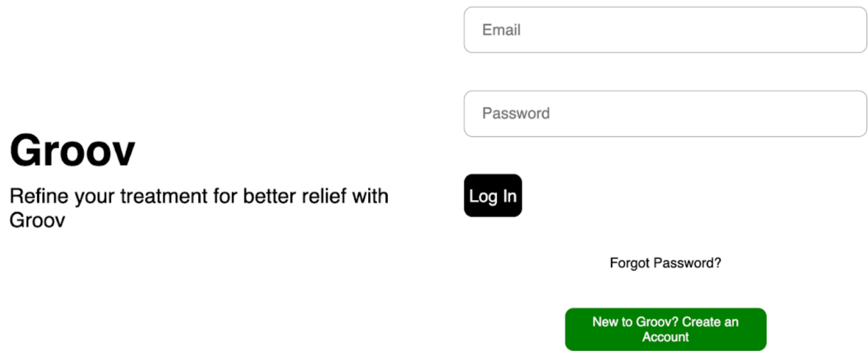
Since last semester, we have expanded Groov into a multi-faceted platform that includes a journaling system through the user front end, a recommendation system for specific strains and treatment types via the retail dashboard, and a social media platform for medical marijuana users to learn from one another. The front-end development has evolved into a comprehensive application catering to both medical marijuana users and dispensary pharmacists, while the UX design has been updated to enable Groov to function as a social platform.

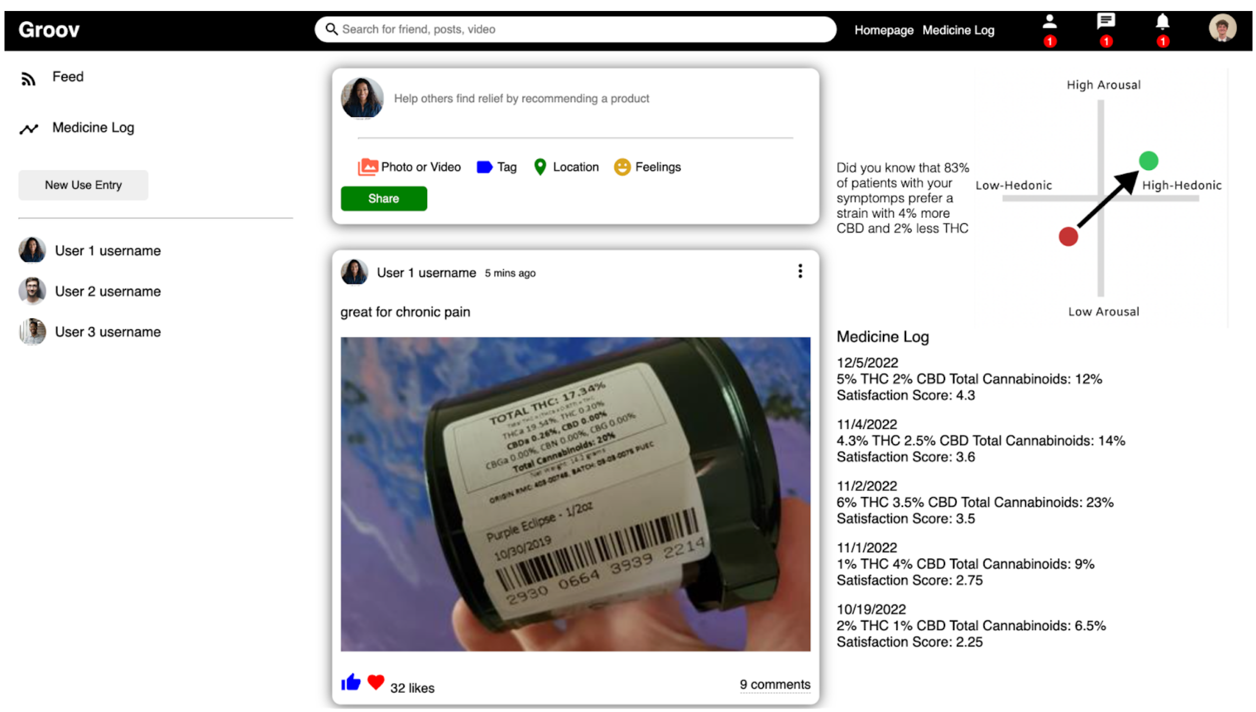
On the data side, our team has worked diligently to to create and collect data that will allow us to test the effectiveness of the application, despite the limited patient data available. This has involved developing synthetic datasets and understanding the requirements for testing our platform, with the ultimate goal of improving the its overall utility.

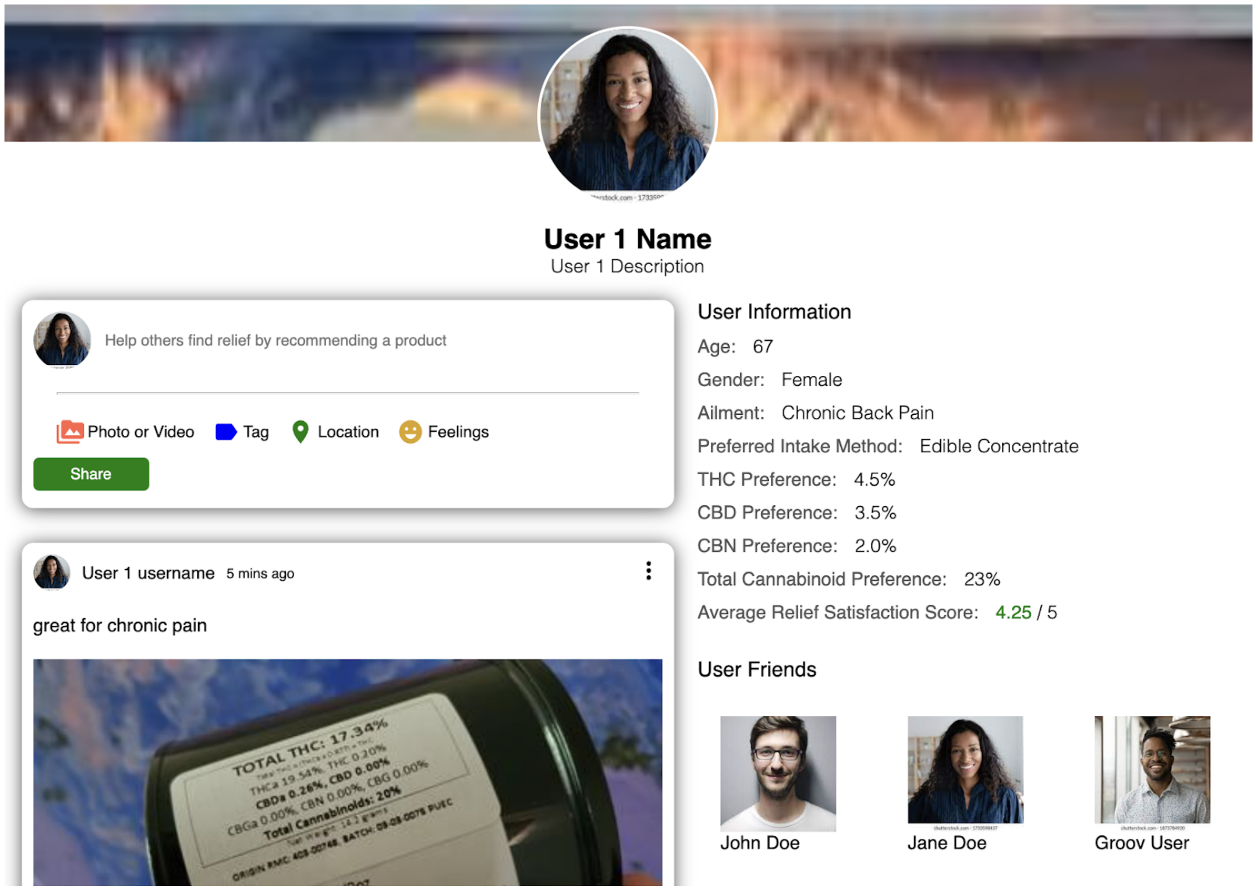
Throughout the development of this platform, we faced several challenges. One of the primary hurdles was creating a reliable data inflow to train our recommendation system. To overcome this, we developed a scheme to test boilerplate learning models on synthetic datasets, generated from limited survey input and guided by existing academic research.

Another significant challenge our team faced was refining our initial idea into a viable final product. The process of transforming a concept into a practical solution required continuous adaptation, learning, and collaboration. We began with a broad understanding of the medical cannabis market, but as we conducted more research we realized the complexity of the ecosystem and the varying needs of different stakeholders. Balancing the requirements of patients, pharmacists, budtenders, and other parties involved in the medical cannabis industry was difficult, and it necessitated numerous iterations and adjustments to our initial concept. Through extensive research, feedback from advisors and industry experts, and in-depth discussions among team members, we were able to narrow down and fine-tune our idea, ultimately developing a more targeted solution. This challenge was a testament to our team's adaptability and served as a valuable learning experience that emphasized the importance of being flexible and receptive to change in product development.

XIII. Appendices:







* Refers to the ABET Course Outcomes 1-7

M&T Portion

Executive summary:

[See page Section II]

Value Proposition:

At its core, Groov is a product that seeks to augment the care experience by providing an assistive tool in the product selection process. The value proposition lies central to that claim, but manifests differently depending on the stakeholder involved. Because this is a product that sits between multiple actors within the system - patients, pharmacists, budtenders - the technical and economic value is fragmented amongst these parties.The value driven is captured in different ways, but this complexity generates challenge when the user might be different than the technical buyer, which could be separate from the economic buyer. Thus, a top-down analysis of the different stakeholders involved will predicate the business analysis.

As for most wellness products, primary value is delivered to the patient. In this case that would be the medical card holder seeking a cannabis product to treat a state-approved condition at the referral of a physician. This stakeholder mainly cares that they receive a quality product in line with the relief needed, and that the recommendation comes from a trustworthy source. The doctor is a similarly aligned party, seeking relief for their patient in need. The nuance with cannabis is that the doctor does not prescribe treatment, or direct any specific product or use - they can only point the patient in the direction of a medical dispensary. The retailer itself is an important stakeholder, and the economic buyer: they care about customer retention and satisfaction, seeking to guide patients to a product that consistently meets their needs and drives predictable sales. Product manufacturers themselves are a stakeholder at the edge of this transaction, creating offerings that meet market demand. They seek a window into how their products are used, and their efficacy with specific patient groups - iterate on R&D with data to complement sales numbers. Finally, cannabis researchers are the other nascent stakeholder at play. As the scientific community opens up to research how cannabis interacts with the body, especially through the medical lens, access to a variety of patient demographics in the clinical setting is quite difficult. States are providing grants for efficacy studies at high rates, and researchers are seeking windows of opportunity. While Groov data might not be collected with a randomized, double-blind methodology - anonymous patient data on how certain cannabinoids affect pain, mood, and energy levels can be quite valuable to augment a study.

Stakeholders:

To understand the core of Groov's value to the retail buyer, we must first look at the role a pharmacist plays in the value chain. Regulations differ state-by-state, but in Pennsylvania each medical dispensary must have a pharmacist on-site to help guide patients in their search for pharmacologically relevant medication. This is important because neither doctors nor salespeople are required any training on the medical science of cannabis. The pharmacist is the lone source of truth for informed guidance, yet time restrains universal access and many patients don't know or don't seek out assistance further than the retail counter. Groov provides value by bringing a baseline toolkit to inform product selection, primarily by mapping a patient's condition to the desired cannabinoid contents for relief - giving the salesperson and patient a dashboard of general attributes they should look for in the product selection process. The value to the retailer is initially guiding a patient to the medically-informed window of offerings, leaving the pharmacist to service higher-order functions such as providing secondary product-specific recommendations, or teaching why products react the way they do. Their own process is improved by patient insights, as the app can show, 'for x% of patients, this chemical combination worked well'. Groov assists the retailer in delivering important insight to the patient in a scalable fashion, and frees up the pharmacist to do their job even better. After talking to multiple dispensaries, managers and pharmacists have validated a willingness to pay for a tool like Groov.

Clear and pertinent value exists to both sides of the point of sale - patients and retailers - yet customer segments diverge in our market research on which retailers to target. Like any seller of consumables, there typically tends to be high and low-end options. From the dispensaries we've researched and visited, our team has found two major customer segments defined on three major attribute vectors: brand, product selection, and price. First, we have our low-end customer - generic branding, contained product selection featuring a handful of recognizable growers, and market-consistent prices higher than the black market but lower than luxury - we deem these, “The Dealers”. Curaleaf, Beyond Hello, and Trulieve would be in this segment. While The Dealers populate most urban areas are market to typical cannabis consumers looking to buy safe, packaged products in a legal environment, the high-end segment often situated in suburban areas such as the Mainline and King of Prussia and markets toward wealthier and first-time cannabis users looking for a wellness focus and a curated retail experience. We call these “The Boutiques” - they typically have cannabis-agnostic branding, a wide product selection of concentrates, topicals, and tinctures along with flower, and luxury-level prices often ~25% higher than market standard. The lack of recreational dispensary offerings means customers looking for high-THC cannabis to get high will seek out The Dealers, while more health and spiritually-conscious buyers will find more value in The Boutiques. While Groov could provide value to all dispensaries due to the lack of training, the value is captured to a higher degree with The Boutique’s patients and the willingness to pay for such a tool is much greater. From this analysis of customer segmentation in the medical dispensary market, our team will focus on the The Boutique segment as a beachhead market.

Market research:
Our team has begun market research through survey of the PA medical cannabis landscape, and through close contact with The Greenbridge Society, a group of cannabis scientists that liaise with growers, dispensaries, and regulators on making cannabis science accessible. Initial regulations permit 150 dispensary locations with more to come, making up our primary customer market. Using other software solutions as a proxy, we would charge an annual licensing fee alongside a nominal cost per customer to each dispensary. Further research is needed to determine pricing, but other revenue streams might provide more fruitful. Greenbridge has given us guidance that growers, researchers, and manufacturers all would have a willingness to pay for patient interaction data. We imagine this to be in the form of a standard quarterly subscription to a Groov Customer Research Report, and further a SaaS fee to interact with our data on-platform. We plan to build out a baseline portal to interact with our collected data, which will serve this function. At this state, that places two major revenue streams: the retail product derived from the value provided in selection at the counter with patients, the data portal becomes a monetization of the patient interaction data used in the app with at-home use. This allows us the flexibility to provide the recommendation and insight value to the patient for free just by using the app.

Competition:
Contained to Pennsylvania, competition in this market is sparse as medical cannabis itself is a nascent market in the state and software platforms that provide secondary value haven’t quite emerged. Tetragram is an experience-logging app that handles the user input side of our value, but doesn’t provide insight directly to dispensaries. They monetize completely off data, and are mostly contained outside the state. The market for medical cannabis itself is small but growing, and the state is nearing movement on recreational legalization. More retailers means a larger market, but the biggest barrier to adoption is education. Many users don’t understand the medical implications of cannabinoid and terpene research, but as knowledge becomes more

accessible this will be a key factor in product choice. When people seek to go beyond “getting high” to seeking out a specific experience, tools like Groov will explode in popularity for their ability to provide second-order insight into the trip users seek. We hope to build the infrastructure to blossom with the PA cannabis market, and help users find their Groov.

Cost

The main expense associated with our project revolve around hosting the server and maintaining the backend infrastructure. These costs include server rental and cloud service fees, data storage and transfer expenses, as well as costs associated with maintaining and updating the server software. Depending on our chosen hosting solution for operating at scale, the costs can vary. Potential solutions include AWS, Google Cloud, and Microsoft Azure with each cloud service providing unique benefits. Moreover, as the platform’s user base grows, additional expenses may arise due to the need for increased server capacity, data storage, and bandwidth requirements. In this case it will be essential to closely monitor these costs and optimize the infrastructure as needed to ensure the platform remains cost-effective and can efficiently scale with the increased demand.

Revenue Model:

Given the unique value proposition and diverse stakeholder landscape, our revenue model aims to create multiple streams of income that cater to the needs and interests of these different groups. By leveraging Groov’s recommendation system, data insights, and analytics, we can establish a sustainable and diversified revenue model that not only supports the platform’s growth but also promotes better outcomes for medical cannabis users and the industry as a whole. The revenue model for Groov will primarily focus on three key areas: subscription services, data insights, and advertising partnerships, all working in harmony to create a sustainable and diversified income stream for the platform.

To start, subscription services will form the foundation of our revenue model. Groov will offer a tiered subscription plan for dispensaries, particularly targeting the high-end boutique dispensary space. The basic subscription tier will grant access to the recommendation system, enabling dispensaries to provide personalized strain suggestions to patients. As dispensaries see the value in the platform, they can upgrade to higher tiers, which will offer additional features and benefits, such as in-depth analytics, enhanced support and access to exclusive research insights.

Another crucial component of the revenue model is the sale of data insights. The platform will gather valuable, anonymized data on user preferences, strain efficacy, and consumption patterns. This data will be highly valuable to various industry stakeholders, such as product manufacturers, growers, and researchers. By packaging and selling these insights, we can help businesses make informed decisions, optimize their product offerings, and contribute to the overall advancement of medical cannabis research.

Lastly, advertising partnerships will serve as an additional revenue stream for the platform. By collaborating with trusted industry partners, such as product manufacturers and ancillary service providers, Groov can offer targeted advertising opportunities on the platform. These partners will benefit from increased exposure to a highly engaged audience of medical cannabis users, while Groov can generate advertising revenue without compromising the user experience.

By combining these three revenue streams — subscription services, data insights, and advertising partnerships — we can create a robust and sustainable revenue model that caters to the needs and interests of various stakeholders in the medical cannabis industry. This approach will not only support the platform's growth but also foster better outcomes for medical cannabis users and the industry as a whole.