Team members:
- Kasha Muzila
- Karl Eirich
- Robert Turnage
- Cynthia Zhu

“I can’t plant more than 75% of them!”

“I spent all that time and money just to wait for a different time to plant!”

The Team

Kasha

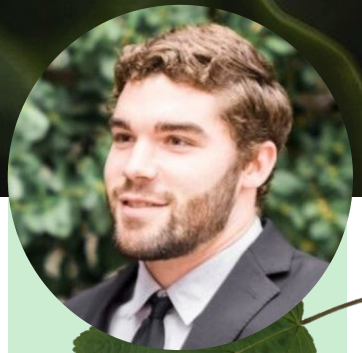
Machine Learning Engineer & Data Engineering

Robert

Product Manager & Information Architecture

Cynthia

Data Science & Machine Learning Engineer



Karl

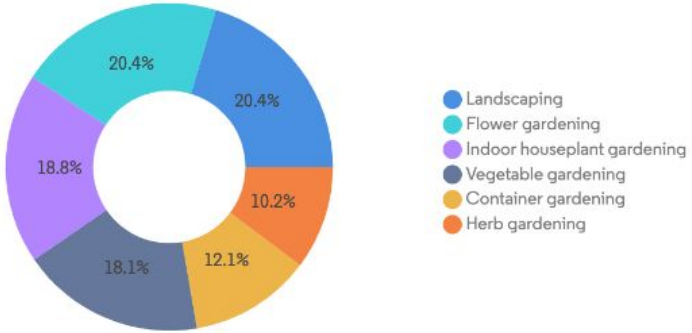
Data Architecture & Engineering

Website and mobile app that can suggest what to plant in a hobby garden based on the number of sun hours, zone, time of year, the direction your garden is facing, and frost dates

- Recommendation Technology
- Simplify the learning curve for beginner gardeners
- Make budgeting for advanced gardeners more efficient

Market Opportunity

Global Landscaping and Gardening Service Market: Landscaping and Gardening Participation, United States, 2019



Source: National Gardening Association

$93 Billion

2018

Before Launch:

300 users

After Launch:

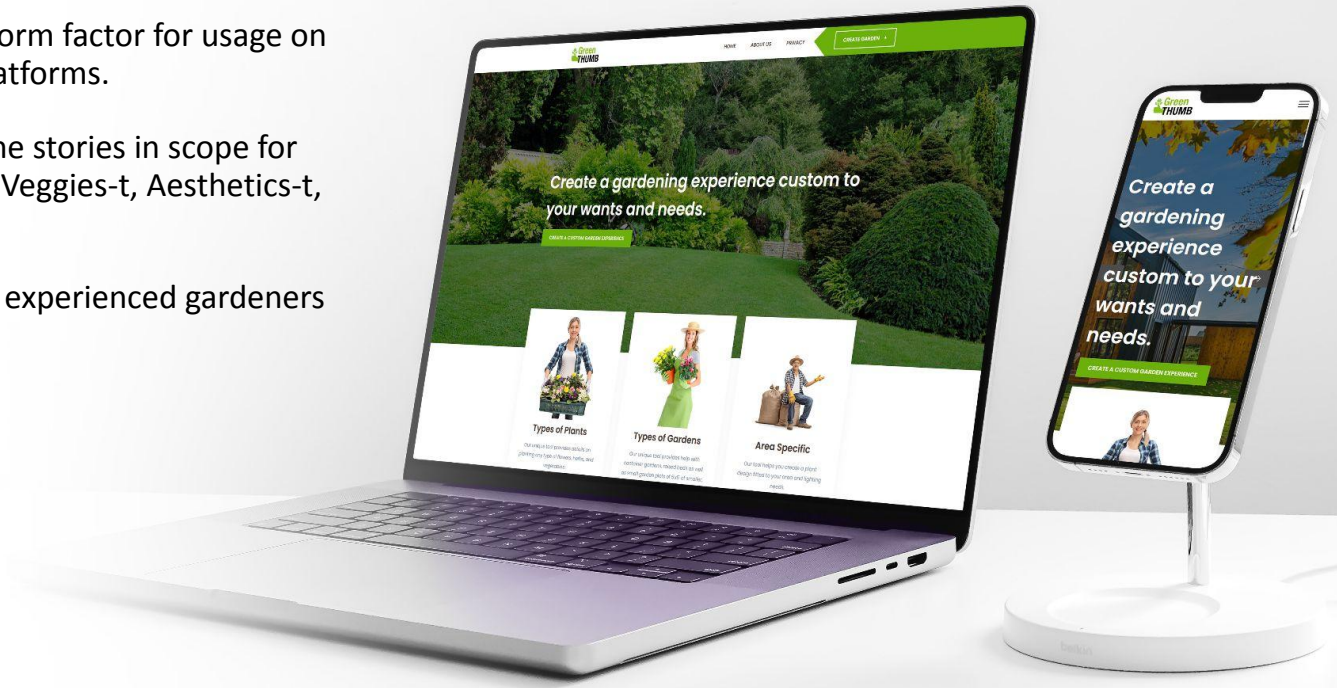
~50,000 users

House Plant Market:

+90,000 users

Let's Take a Look!

- Easy to use having a form factor for usage on Web 2.x or Mobile platforms.
- Design is aligned to the stories in scope for three gardener types Veggies-t, Aesthetics-t, Hybrid-ers.
- Open to beginners or experienced gardeners alike.



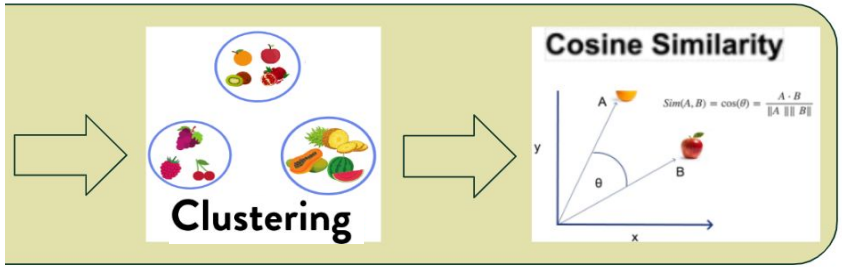
GreenThumb.ai | Homepage

http://20.127.87.137

Data Engineering

- Manually curated data from multiple sources
  - Scraped from companies, universities, and community forums
- Data quality varied between sources
  - Plant pictures were particularly varied
  - Plant details fluctuated in specificity
- Around 50000 plants across 16 unique plant types
  - Data stored in parquet files due to size

Recommendation Baseline



Similarity score

0.3

Metrics

Cosine similarity measures the similarity between plant of interest and suitable plants in the database.

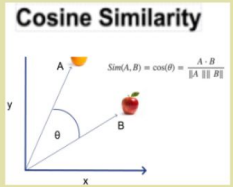
Technical Evaluation

With feature engineering,
43.5% increase in
information entropy

Recommendation Engine

Item-based filtering

Final

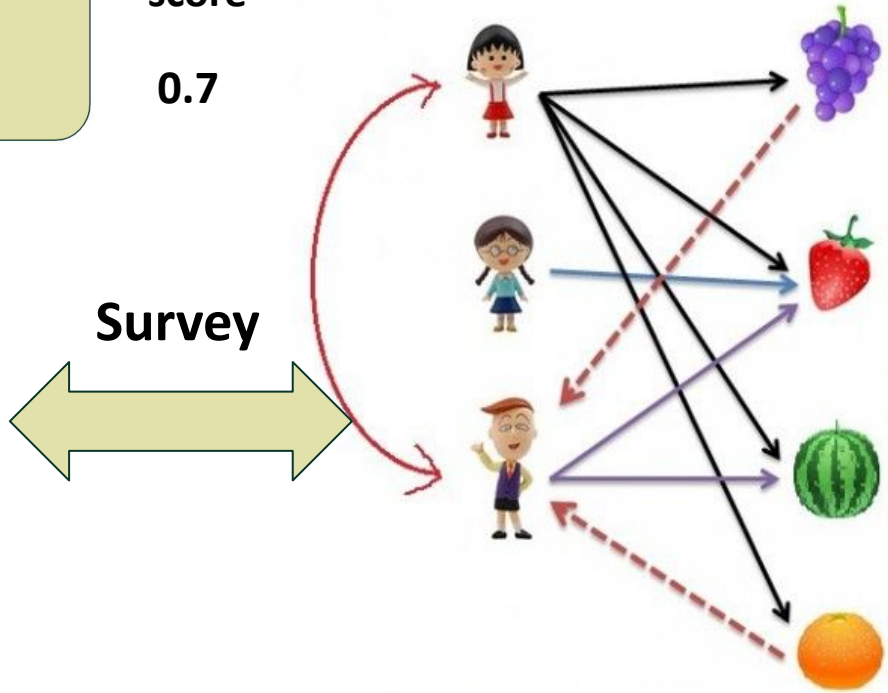


Similarity score

0.7

Survey

User-based filtering



Testimonial: Susan Muzila

"Most Definitely! Because it gives me a chance to see all of the plants so I can narrow it down and give it to our landscaper."

"I like that feature, I bet that there are many websites that don't give you that option at all."

"I love the pictures!...and the added invasive species feature!"

Background

- Age: 55-65
- Married with 2 adult children
- Education: Undergraduate

Role: Executive Assistant

- Job measured: alleviates additional duties of Medical Director
- Reports to: Medical Director
- Skills required: organization, communication, project management

Company Information

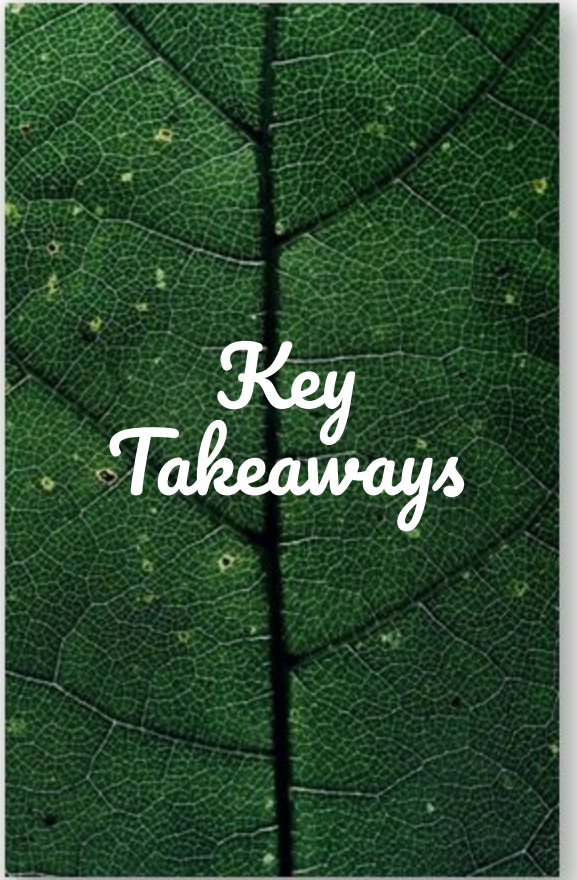
- Industry: Medical
- Yearly Revenue: $60k
- Employees: 5000

[redacted]

Try Green Thumb AI Today!

[redacted]

https://youtu.be/no4zo4ZibeM



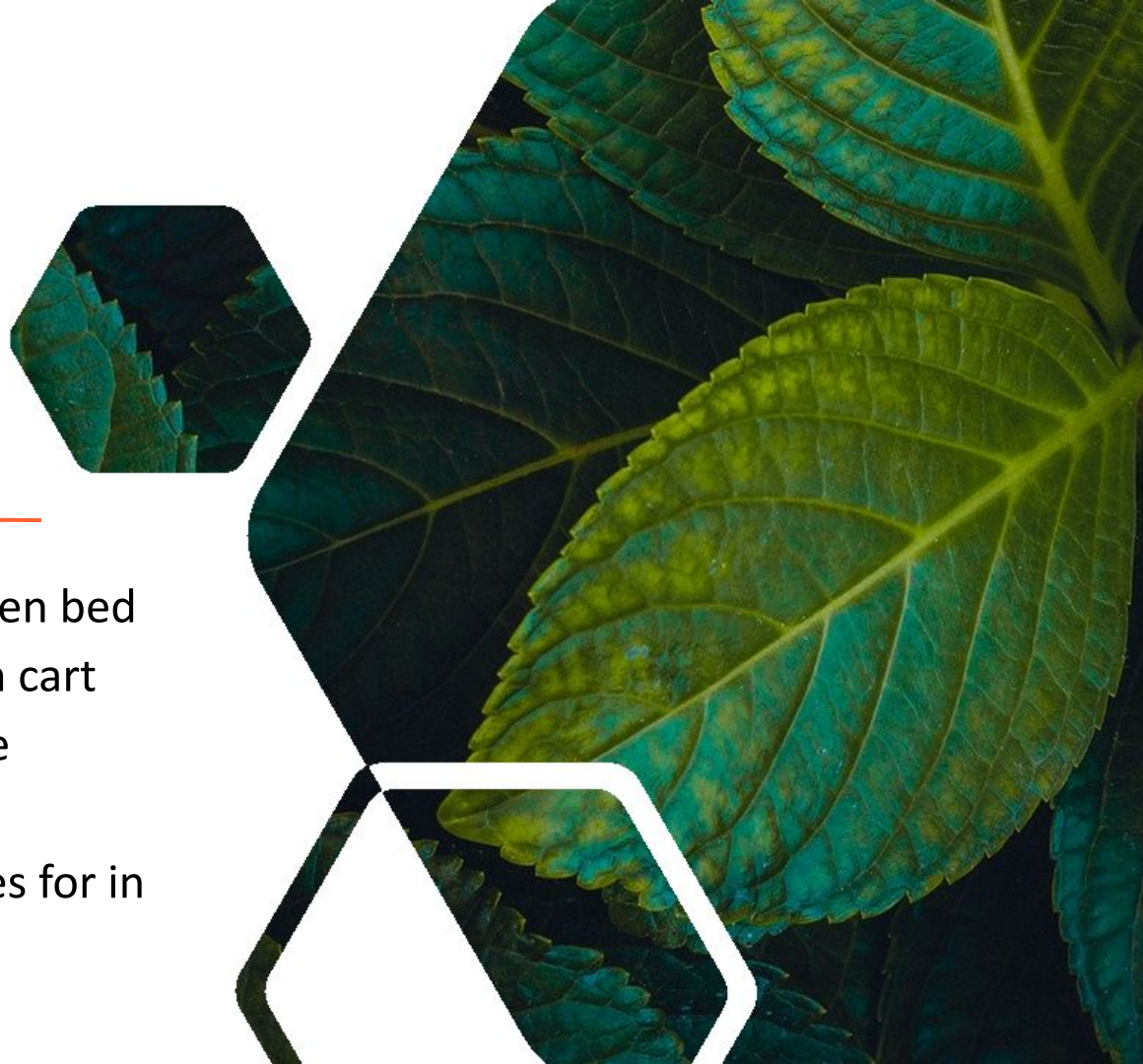
3 Technical Challenges

- Scrapping the initial plant dataset
- Unifying and normalizing the plant dataset
- Huge learning curve with building the front end of the website

3 Detailed Learnings

- Effective labeling had a huge improvement on modeling
- Using K-means clustering is computationally unstable with unique dataset at each run.
- Using Global Horizontal Irradiation (GHI) instead of Direct Horizontal Irradiation (DHI) increased range in recommendations

Future Opportunities



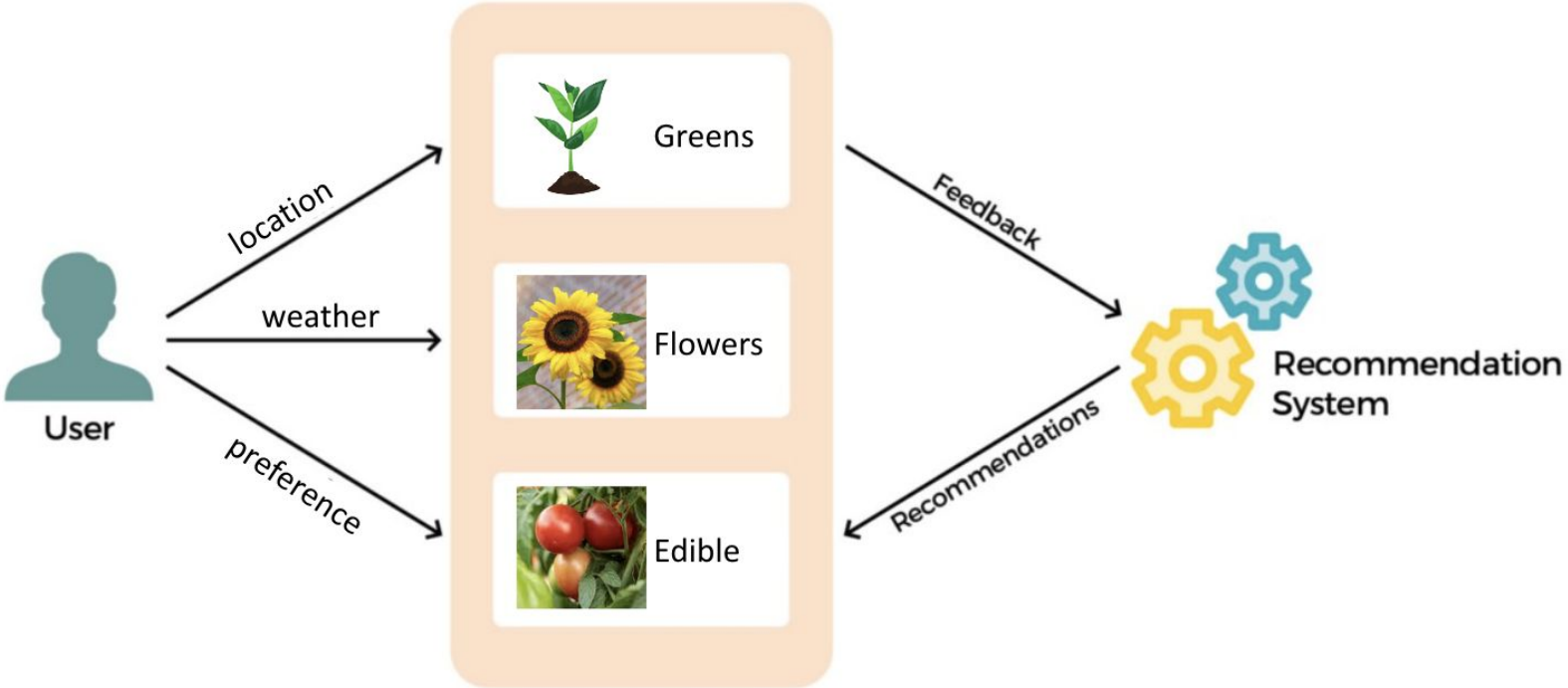
- Machine visual to assess garden bed
- Suggest plant layout based on cart
- Connect to partner for co-sale opportunity
- Funnel user to local businesses for in person shopping experiences

Wrap Up

Our app has a data science algorithm that provides suggestions on what to plant in the hobby garden based on initial environmental conditions to ensure that anyone interested in starting a hobby garden retains their excitement and has a fun and easy experience while planning.


Thank
You

Recommendation Engine



Overview

Kasha

Problem (Kasha's Story)

While turning my apartment balcony into a semi-self-sustaining vegetable garden, I have run into the bottleneck of figuring out what to plant based on the initial environmental conditions of my balcony. I have spent dozens of hours researching to finally come up with a list of vegetables to plant, only to find out that I cannot plant half of them because it is not during the optimal season. This also includes pest-resistant plants that are an imperative companion plant for some vegetables, causing the number of feasible vegetables to plant to be less than half of everything purchased.

Impact

Why is this problem important to solve?

A website/app that can suggest what to plant in a garden based on the number of sun hours, zone, time of year, the direction your garden is facing, and frost dates would help save time and money when planting a garden. Additionally, the learning curve for beginner gardeners is so high that many are too afraid to start. This website/app would be able to simplify the learning curve into smaller chunks so beginners can purchase and learn about the plants that they can plant right now instead of all plants from every season. For more advanced gardeners, this would make budgeting more efficient as they can see what they can plant based on their initial preferences and can price compare before investing. A future outlook for this project is to function like the Kayak website but with consolidating gardening products into one website based on the user's preferences. These preferences can be specific. For example, some users do not like to purchase from Amazon and prefer to purchase from a small business.

Green Thumb AI is a website and mobile app t
suggest what to plant in a hobby garden based on
number of sun hours, zone, time of year, the direct
garden is facing, and frost dates.

Additionally, the learning curve for beginner garde
high that many are too afraid to start. This app wo
able to simplify the learning curve into smaller chu
beginners can purchase and learn about the plants
can plant right now instead of all plants from every
For more advanced gardeners, this would make bu
more efficient as they can see what they can plant
their initial preferences and can price compare bef
investing.

All of this would help enable gardeners of varying s
knowledge to save time and money.

Problem Statement

Kasha

- Improving the bottleneck of determining what to plant based on the initial environmental conditions
- Website/app would be able to simplify the learning curve for beginners.
- Would make budgeting more efficient for advanced gardeners

Future product vision (Karl)

1. Using vision recognition assess garden bed area
2. Using python recommend layout based on shopping cart selection of diameters to meet possible geometries for aesthetics or functional companion outcomes.
3. BCOM connection with partners for co-sell opportunities driving platform of community of market share. Similar business models to Streaming Music or eBooks. Eg. take shopping cart to Home Depot and check out via API purchasing.
4. Provide online shopping experience for BCom to lesser business entities like local nurseries.

** 5. Allow user to provide ratings/feedback

How can your model approach and way of solving the problem be generalizable?

- Potentially generalizable to other similar or adjacent problems

Market Research (Kasha)

1. Market research, extension, wider products, etc…
    a. Market size
    b. User size
    c. Quantification on potential level of impact that product can deliver now and in the future if continued to work on it
        i. Educated guess
2. How many items are in this universe, how will the DS keep up?
3. Hobby gardening promotes mental health, which extends the use case
4. Show how large plant data is in comparison

A future outlook for this project is to function like the Kayak website but with consolidating gardening products into one website based on the user's preferences. These preferences can be specific. For example, some users do not like to purchase from Amazon and prefer to purchase from a small business.

1. Presentation 3 Delivered on Dec 5th
    a. Fill out this form: https://www.ischool.berkeley.edu/programs/mids/capstone/2022b-summer
        i. Add demo video to this page
    b. Things to add
        i. Show App, Data, etc at scale architecture design
        ii. Show Market space interaction where plant vendors can join and interact with gardeners for market share and influence

Wrap-up (Karl)

1. Put sentence that states mission and approach (aka elevator pitch minus the team introduction and the ask)

Ever struggled to learn gardening, and always wanted to learn but don't have time. Our app has a data science algorithm that provides suggestions on what to plant in the hobby garden based on initial environmental conditions. Now the budding novice can become the master with guidance and ease, no waste and desired results.

My team consists of four FTEs has a go to market strategy with significant market opportunity. First to data and customers wins the land grab!

How about a follow up meeting to take a deeper dive into the product feasibility as it is today and what we know about the opportunity.

Key Takeaways (Kasha)

1. Top 3 technical challenges you have overcome
2. Top 3 key detailed learnings from the model evaluation and error analysis

Feedback

1. Get user feedback on usage experience
2. get ground truth baseline from users experience for retrain cycles

Robert

UI/UX

Having a compliant form factor for user usage on web 2.x or Mobile platforms.

Design is aligned to the stories in scope for three gardener types Veggies-t, Aesthetics-t, Hybrid-ers.

Still to come,... the design for user inputs that drive the model and the usage of outputs. Dependent on what we can achieve in time…

GreenThumb.ai | Homepage
http://20.127.87.137

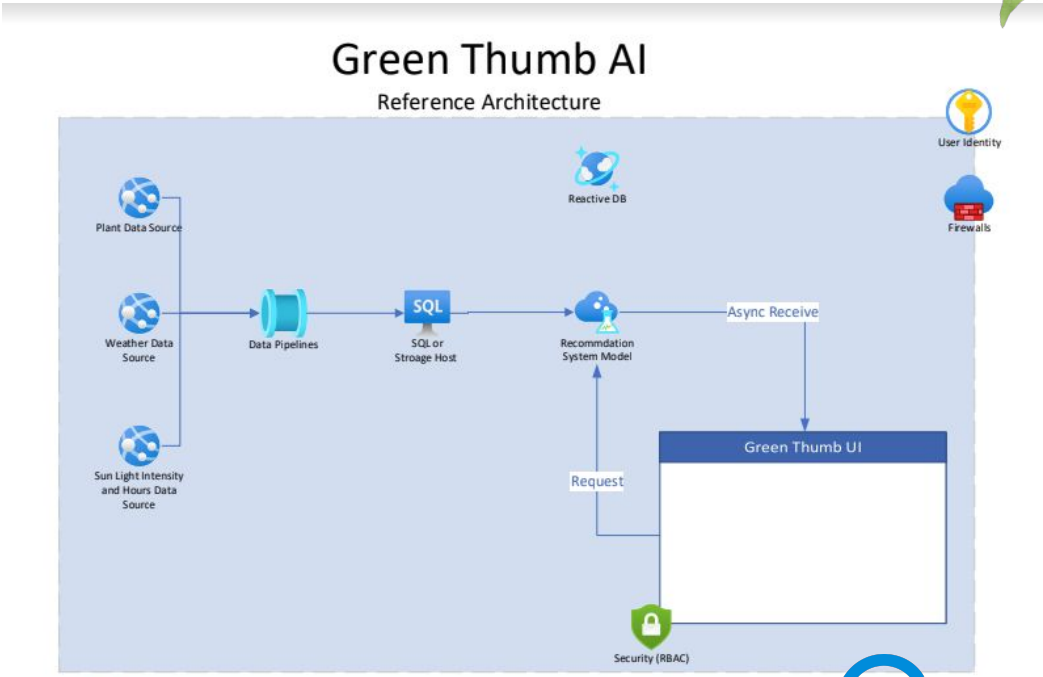
Robert

Architecture

This is a simple diagram showing our reference architecture.

Data ingestion via pipelines in python 3x. Clean and initial transformation.

Final transformation and data science recommendation system pairing with MLOps on Cloud Machine Learning Platform (AML)

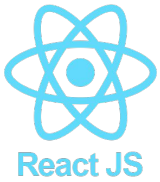


Berkeley
UNIVERSITY OF CALIFORNIA

Robert

App Engineering

Azure Machine Learning



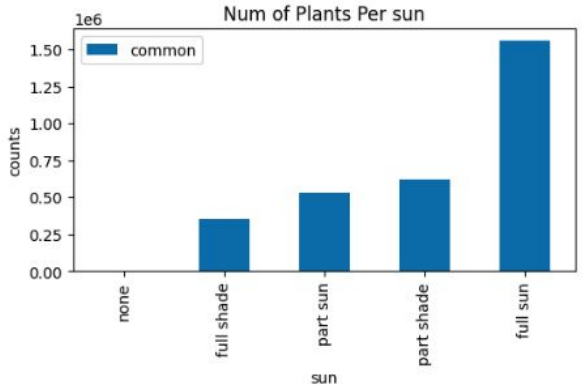
We are using HTML5, CSS3, JS (Bootstrap, React):

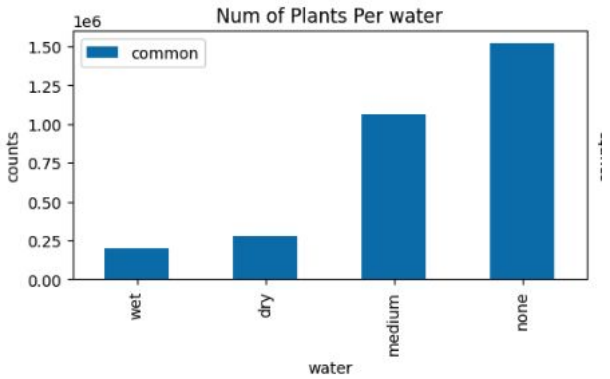
Hosted on Azure VM
Ubuntu 20.04 LTS
Using in Python 3.x
Flask 2.x

With the logic for the data science model trained and hosted on Azure Machine Learning

Data Engineering

- Data normalization was done to increase join compatibility
  - The process was very manual but provided the granular control needed to align value formats
  - Columns were dropped that did not see data parity across other tables
  - Values were aggregated to prioritize the join without compromising the signal



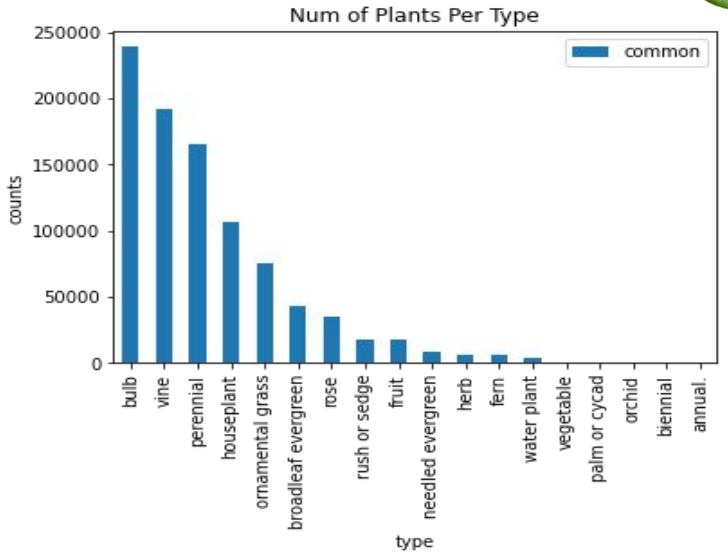


EDA - plants

Total plants: 3244, if excluding shrub/tree, 1143.
All Features: 25 features
-'attracts', 'flower', 'fruit', 'leaf', 'water', 'genus', 'height', 'width', 'special_feature', 'scientific_name', 'description', 'soil_conditions', 'texture', 'emergence_days', 'maintenance', 'native_region', 'propagation', 'bloom_time', 'sun', 'problem_solvers', 'type', 'zones', 'family', 'color', 'temp_bucket'

Feature Engineering:

- Explode each category of feature into a separate record
- Bucketize temperature to match the lookup table
- Type of plants
- Check features distribution
  - 10 Features have >50% none value: Attracts, flower, fruit, leaf, scientific_name, description, soil_conditions, texture, native_region, emergence_days

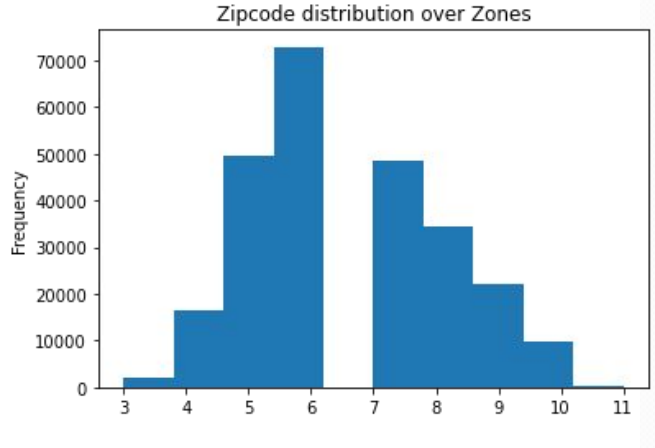


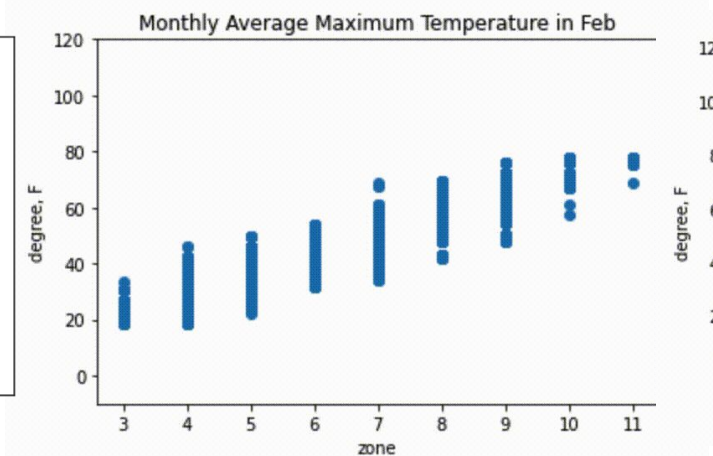
	attracts	common	flower	fruit	leaf	water	genus	height	width	special_feature	scientific_name	description	soil_conditions	texture
0	['butterflies']	chinese abelia	['showy', 'fragrant']	['none']	['none']	['medium']	abelia chinensis	[5]	[3]	['hedge']	none	none	['none']	['none']
1	['none']	abelia	['showy']	['none']	['good fall']	['medium']	abelia 'edward goucher'	[3]	[3]	['hedge']	none	none	['none']	['none']

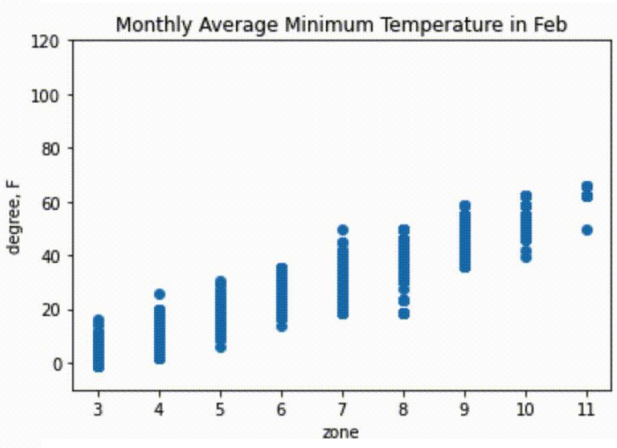
temperature	emergence_days	maintenance	native_region	propagation	bloom_time	sun	problem_solvers	type	zones	family	color
[0]	[0]	['low']	['china']	['none']	['july', 'september']	['full sun', 'part shade']	['rabbit', 'deer']	['deciduous shrub']	[7, 8]	caprifoliaceae	['white']
[0]	[0]	['low']	['none']	['none']	['may', 'september']	['full sun', 'part shade']	['erosion']	['deciduous shrub']	[6, 7, 8]	caprifoliaceae	['lavender']

EDA - location-based weather dataset

- Joined weather with planting zone by closest geospatial coordinates
  - 210 cities with 2017-2020 temperature vs 21K zip code with designated planting zone → most of zip codes will share the same weather features
- Zone has a relative normal distribution across US.
- Zone correlates with temperature over Jan - Dec

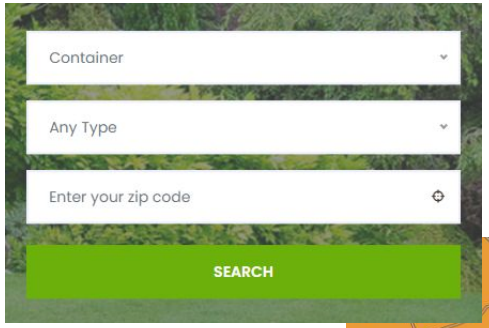




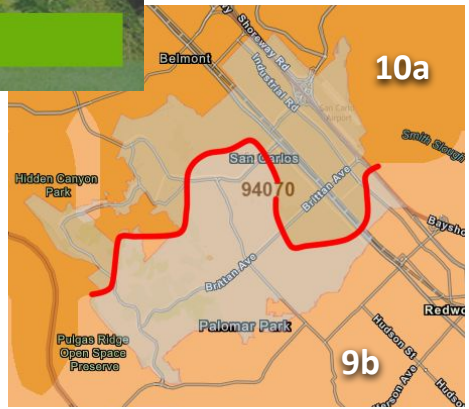


Model

1. Filter Plant Dataset Using Zip Code



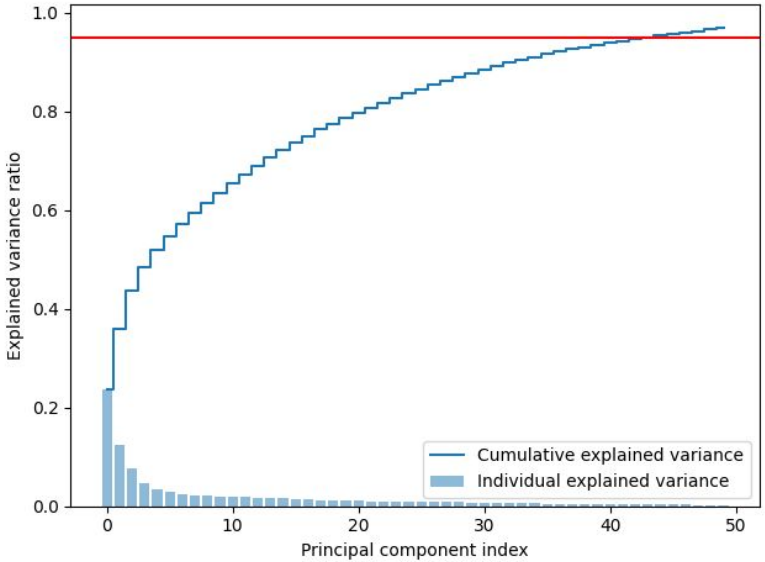
Filter



2. Use DS Model to Recommend Best Plants to Grow

Model

- Given the user's zip code, we find zone, sunhour, temp_bucket based on tmax and tmin → subset plant data that meet above criteria
- Remove the features have >50% null value
- PCA and selected cutoff at 95% explained ratio
- Modeling:
  - DBSCAN to select optimum number of clusters → all -1 label, no proper cluster found
  - KMeans, gridsearch give k=300 with silhouette score of 0.18.



Data Privacy Consideration Overview

- *Team Name: Green Thumb AI*

  Description: Algorithm that suggests recommendations for what to plant in the small garden based on initial environmental conditions.

Data Privacy Consideration	In-Scope for Capstone Project to address. Actions:	Out-Scope for Capstone Project to address. Possible solution:
FTC Principles circa 2012	Data Collection (Zip Code). We do not persist this data and we do not aggregate or persist with IP tracing or tracking. Data Usage is within the session state and is ephemeral.	Security considerations, should be designed and implemented for production. Current state of system can implement B2C/B2B RBAC and identity management for AuthN/AuthZ requirements.
CALOPPA	There is no tracking via web or mobile. Azure platform is used for hosting in the cloud Data Privacy in the Trusted Cloud | Microsoft Azure	(BCOM) PCI-DSSv3 will not be enabled in this feasbillity version, but is a consideration for production ready edition.

Next Steps .

Post MVP


Karl
Thank
You