Tree of Life Counseling Mental Health Matters: Improving Access to Resources

12/06/2025

AI Studio Final Project

BREAK THROUGH TECH

Meet Our Team

Alvisa Krasniqi
Case Western Reserve University

Hakim Fessuh
University of Maryland, Baltimore County

Blessing Ogunfowora
The University of Texas At Dallas

Naman Bagga
Rutgers University

Our AI Studio Coach & Challenge Advisor

Zach Sloan
Tree of Life Counseling

Tianna Ramos
Tree of Life Counseling

Haziel Andrade Ayala
Cornell Tech

Presentation Agenda

Introduction & Project Overview	4 minutes
Data Understanding & Data Preparation	4 minutes
Modeling & Evaluation	4 minutes
Findings & Future Work	4 minutes
Final Thoughts	4 minutes

AI Studio Project Overview

The project aims to help Tree of Life Counseling target its marketing efforts more effectively, increase morning bookings, and enhance operational efficiency by predicting which types of clients are most likely to book morning appointments (9–12 p.m.)

Business Overview

- Tree of Life Counseling Center is a family-owned mental health practice that provides both therapy and medication management services.

- As a for-profit private practice, Tree of Life's mission centers on promoting accessible, high-quality mental health care and improving overall client well-being.

Our Goal

1. Analyze client demographics to uncover patterns related to appointment times across Tree of Life Counseling's locations.
2. Identify which clients are more likely to book for hard-to-fill morning appointments (9–12 a.m.).
3. Apply supervised learning models, with an emphasis on multi-label classification, to predict and understand these relationships.
4. Provide actionable, data-driven insights that help Tree of Life Counseling improve marketing strategies, increase appointment bookings, and enhance operational efficiency.

Business Impact

1. Increase appointment bookings, especially for underfilled morning time slots.

2. Optimize marketing strategies.

3. Improve operational efficiency through data-driven decision-making on client targeting and scheduling.

Our Approach

Sep 2025 - Oct 2025

Data Exploration and Preparation

Load and clean Tree of Life Counseling’s CRM data (≈1,900 records). Handle missing values, outliers, and categorical variables through one-hot encoding and normalization.

Oct 2025

Exploratory Data Analysis (EDA)

Transform categorical variables into predictive features. Apply supervised learning models, particularly multi-label classification, to predict which referral sources are most likely to produce clients for morning appointments.

.

Oct 2025 - Nov 2025

Feature Engineering and Modeling

Applied supervised learning models, focusing on multi-label classification, to predict what customers are more likely to book morning appointments.

Nov 2025 - Dec 2025

Model Evaluation and Optimization

Measure performance using metrics like accuracy, precision, and F1-score. Refine the model through hyperparameter tuning and validation.

Resources We Leveraged

1. BTT ML Foundations Course and Python ML libraries - (Scikit Learn, NumPy, Pandas, MatPlotLib, Seaborn, Folium, GeoPy)
2. Challenge Advisors (Allison D'Mello & Zach Sloan) & AI Studio Coach (Haziel Andrade Ayala) Bi-weekly meetings and communicated any roadblocks
3. Google Colab - Development
4. Github Projects - Track Tasks
5. Slack & Whatsapp - Team and advisor communication

Summary of Insights & Key Findings

1. Morning appointments are the hardest to fill, confirming the initial business challenge.

2. After preprocessing, 1,600+ usable entries remained, indicating successful cleaning and filtering.

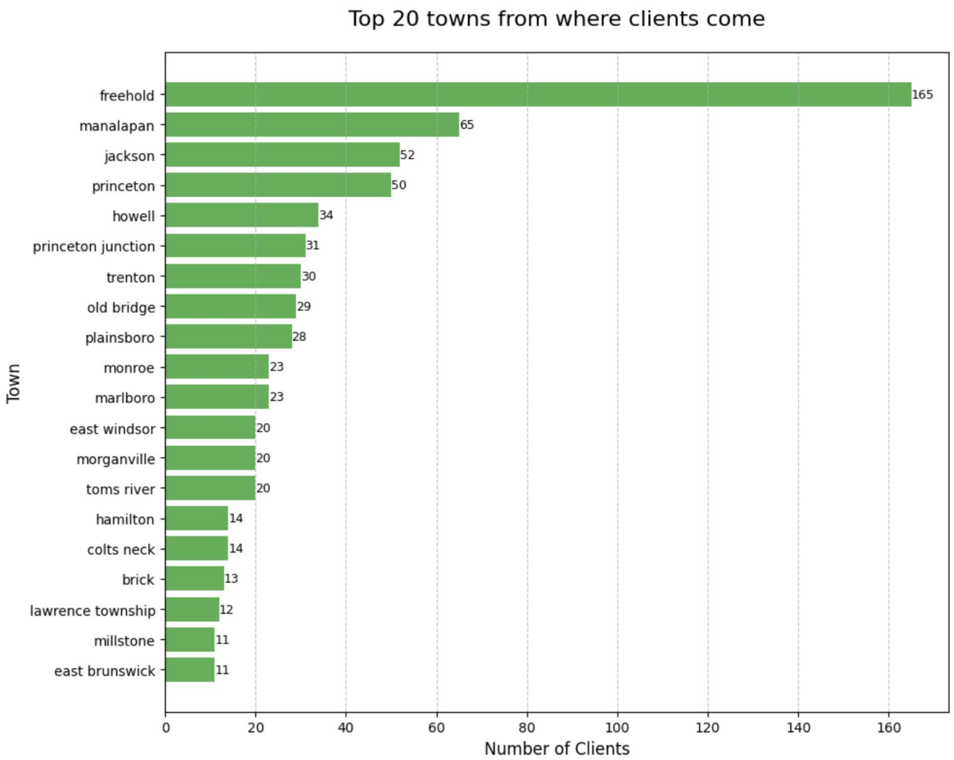
3. Preliminary EDA setup (e.g., visualizing distributions with Matplotlib and Seaborn) confirms there are no major anomalies or outliers affecting the dataset.

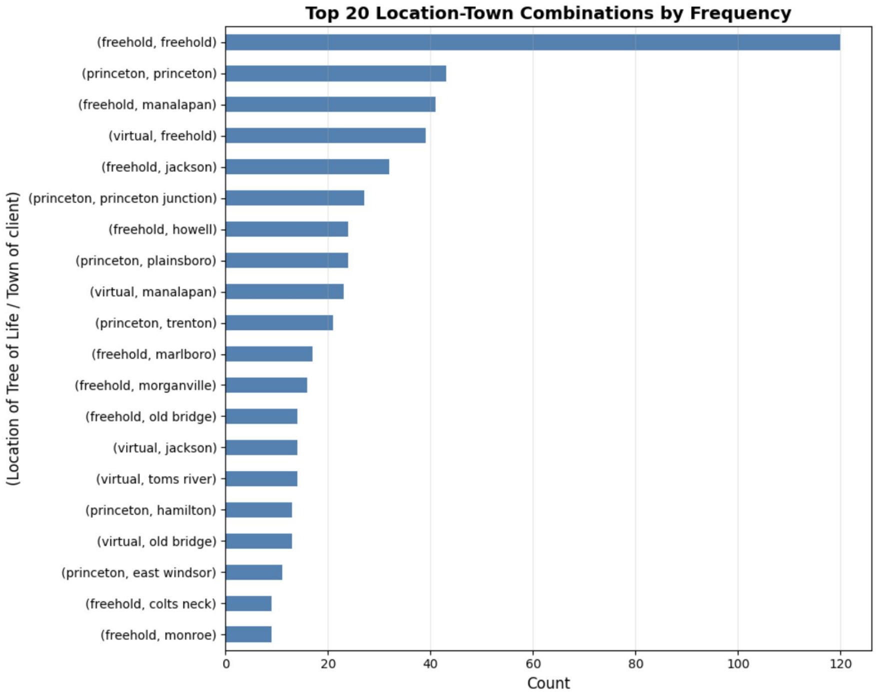
Data Understanding and Data Preparation

Data Understanding

1. Data Source: from internal databases of Tree of Life Counseling
2. Data contains:
    a. 1964 records and 17 features
    b. Range of missing values among columns: [21, 1733]
    c. 3 features have more than 1300 missing values
    d. Inconsistency in data input:
        i. appt_type is of form: morning (9-12), 09:00 am, etc.
        ii. for couple therapy, age is mentioned as ‘45-50; 50-55’ which represents 2 people
    e. 415 records have missing label

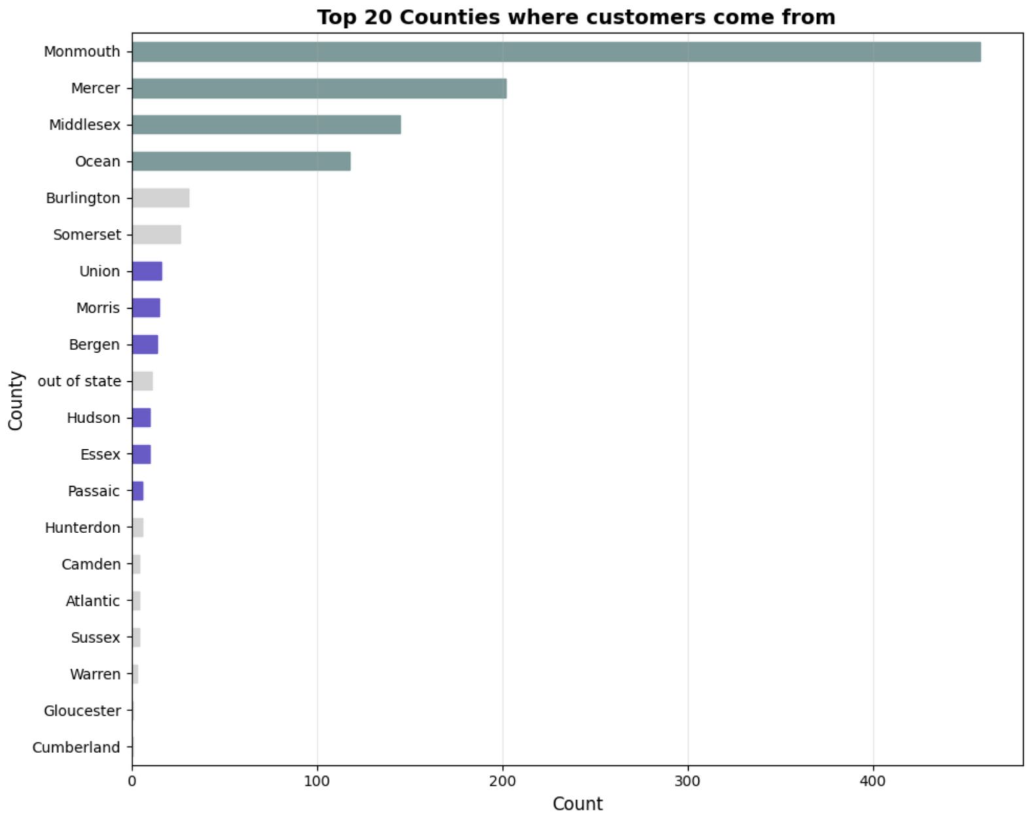
KPI: New Location of Tree of Life





These towns give alot of customers. **Observation**: Freehold, Princeton clients prefer local location. Whereas, for the clients from other location, virtual visit is also a very popular option.

KPI: New Location of Tree of Life

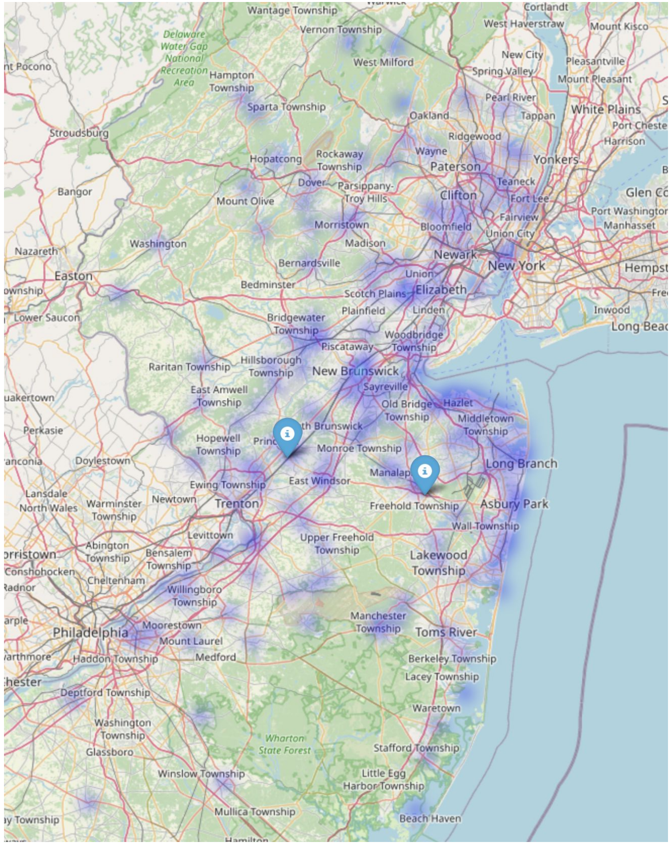


The top 20 towns from where customers come are from only 4 counties: Monmouth, Mercer, Middlesex, Ocean (in pale blue)

Which means, customer base in other counties is very low. However, other potential location cant be decided based on most customers. We need to explore customer density.

Since, the counties highlighted in purple, forms a cluster around each other, and also offers a wide customer base together, complemented with NY City.

KPI: New Location and Billboard



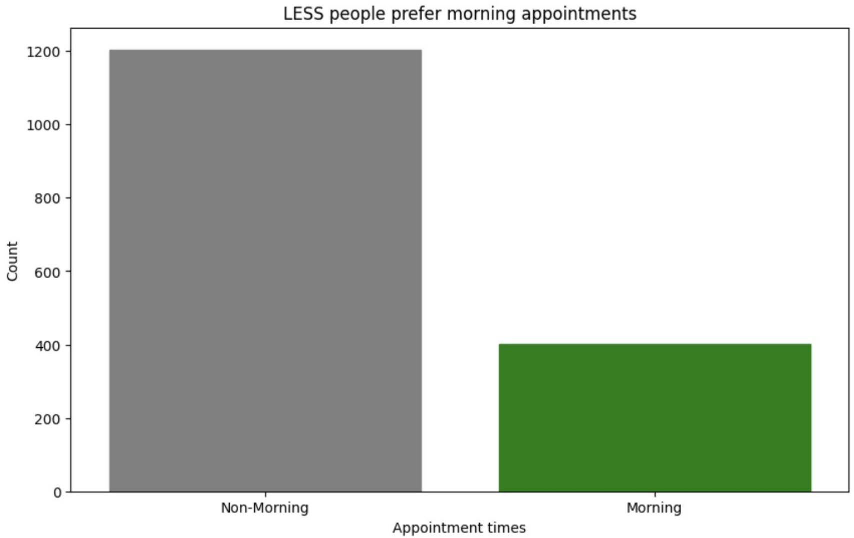
Heatmap of customers in NJ plotted using Folium and Geopy. Pinpoints are 2 TOL locations, serving Central and South NJ.

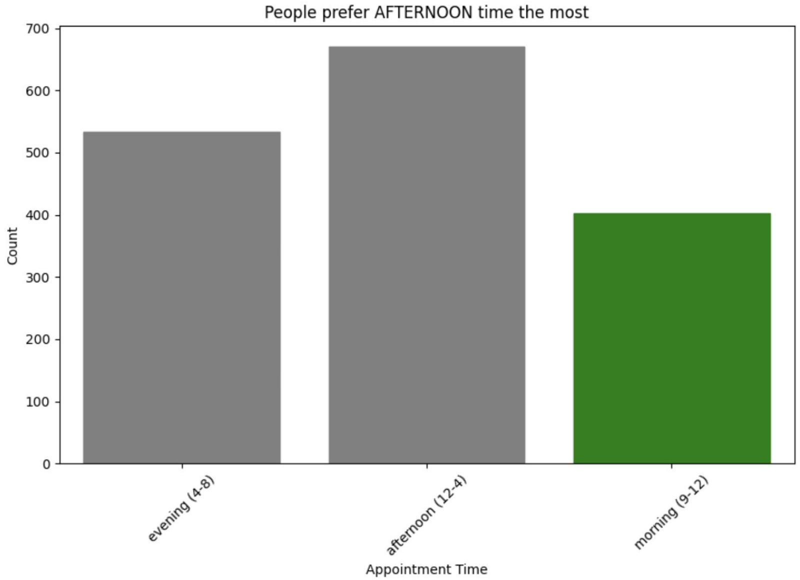
New Location: If we focus on **North NJ**, **Newark** could serve as a potential location or any metro town in Hudson or Union county.

Billboard: Since it was a dream of TOL family to be on a Billboard, **US1 and US9** could be the suitable highways where you can advertise TOL.

Use this interactive map to understand the geographics of the customers, which will help you to decide better about potential locations or potential advertisement locations. You can find this map at: https://drive.google.com/file/d/1vetxX2QvDTD-Oli3t5HZO5sp22sEuboS/view?usp=sharing

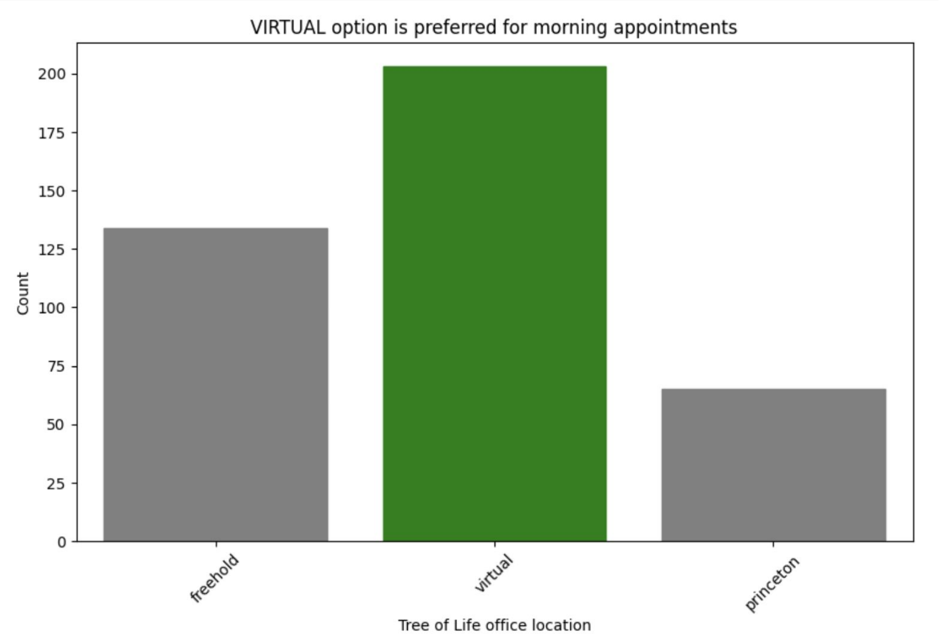
KPI: Potential morning customers





It is clear, morning appointments are least popular

KPI: Potential morning customers



People prefer virtual appointments for morning

Modeling and Evaluation

Model Evaluation

- Models Trained: Logistic Regression, GDBT, Random Forest, KNN
- Target Label - is_morning = 1, non_morning= 0
- Dataset Size ~1600 cleaning records
- Total Features -244
- Feature examples - client age range, Therapy Type, Insurance plan, office location, and Intake method

Model Comparison

Model Name	Description	Results	Pros	Cons
Logistic Regression	Linear model that predicts whether an appt is morning or non-morning based of our weighted features	• Accuracy: 0.75 • ROC AUC: 0.67 • Recall (Morning): 0.09	• Simple and fast to train • Clearly shows feature influence (positive/negative impact	• Performed poorly on imbalance data • Fails to capture complex non-linear relationships
GDBT	Builds many small trees in steps — each fixes the previous one's mistakes.	• Accuracy: 0.75 • 0.67 • Recall (Morning): 0.09	• Captures complex patterns • Shows feature importance • Performs well on structured data	• Biased toward non-morning class • Needs more tuning

What KNN Learned about Morning Bookings

- Age range- Some age groups showed higher morning demand
- Referral Source - Physician and clinical referrals leaned towards mornings
- Office Location- Certain locations had more morning bookings
- Intake Method - Phone intake aligned with morning sessions

Key Insight:

- KNN Recovered 25% of morning sessions vs ~9% from other models.

Challenges and Solutions

- **Challenges:**

1. Imbalanced labels
2. Low recall scores
3. Small dataset

- **Solutions:**

1. Used a simpler model
2. Compared results across four models
3. Optimized for recall

We needed more relevant information such as gender, exact age, occupation (such as student, housewife, etc.)

Final Thoughts

What We Learned

- Gained hands-on experience with pandas, NumPy, Matplotlib, and Seaborn for data cleaning and visualization.
- Implemented one-hot encoding and improved data preprocessing and teamwork using GitHub and Colab.
- With more time, we'd focus on model training and validation to refine marketing insights.

Thank you!

Questions?