

Kenya Tree Planting Analysis

Hotspots, Gaps & Seasonal Alignment

A Data-Driven Approach to Equitable Reforestation

CAPSTONE PROJECT

DENNIS KIPYEGON

Problem Statement & Business Objectives



Problem Statement

1 Kenya has launched ambitious reforestation initiatives, but
there is limited visibility into:

2 Where planting is happening

Whether it aligns with favorable ecological conditions

If efforts are equitably distributed across counties

Without spatial and temporal insights, there's a risk of:

- Over-concentration in accessible areas
- Planting during dry seasons (low survival)
- Under-serving arid or marginalized regions



Business Objectives



Map planting hotspots and gaps to guide resource allocation



Assess alignment with rainy seasons to improve survival rates



Benchmark counties by equity (trees per capita), not just volume



Provide decision support for Kenya Forest Service and county governments

Data Understanding & Sources

county_tree_planting_summary.csv

- ✓ county_tree_planting_summary.csv
- ✓ County-level planting summaries
- ✓ Aggregated tree planting statistics
- ✓ Population data by county
- ✓ County centroid coordinates 47 counties covered
- ✓ 2019 Census population data

df_gps.csv

- ✓ GPS-tagged planting events with coordinates
- ✓ Date of planting activity
- ✓ Number of seedlings planted
- ✓ Temperature at planting location
- ✓ Rainfall measurements
- ✓ Over 1,200 records across Kenya

Data Preparation & Cleaning Process

GPS String Parsing

Extracted coordinates from GPS strings when lat/lon

```
df_gps['gps_str'] = df_gps['gps_str'].str.strip('"')
```

Kenya Bounds Filtering

Filtered data to Kenya's geographical boundaries

```
kenya_mask = (df_gps['lat'] >= -5.0) & (df_gps['lat'] <= 5.0) &  
(df_gps['lon'] >= 33.0) & (df_gps['lon'] <= 42.0)
```

Outlier Removal

Removed extreme values (>500k seedlings) and missing data

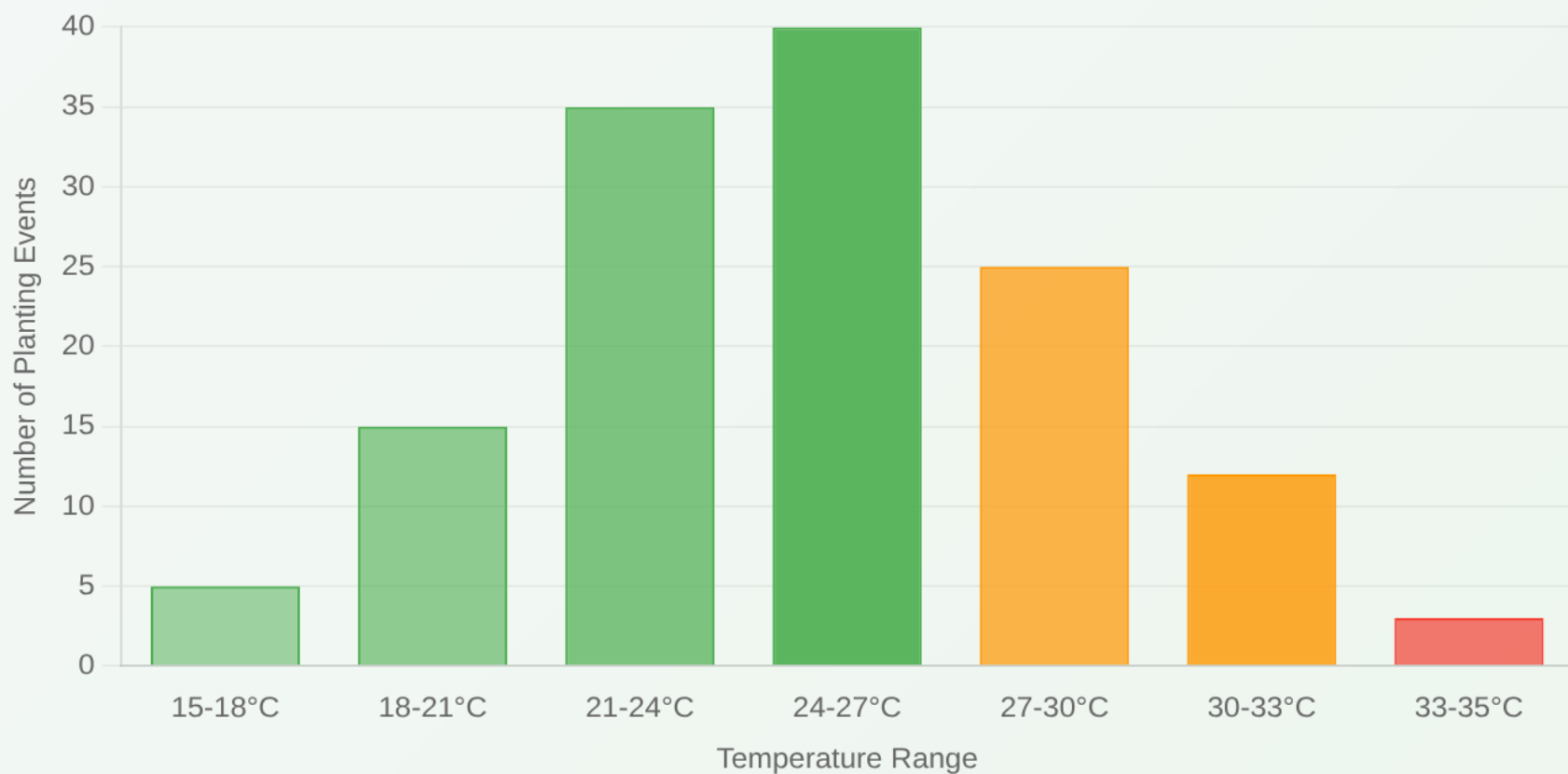
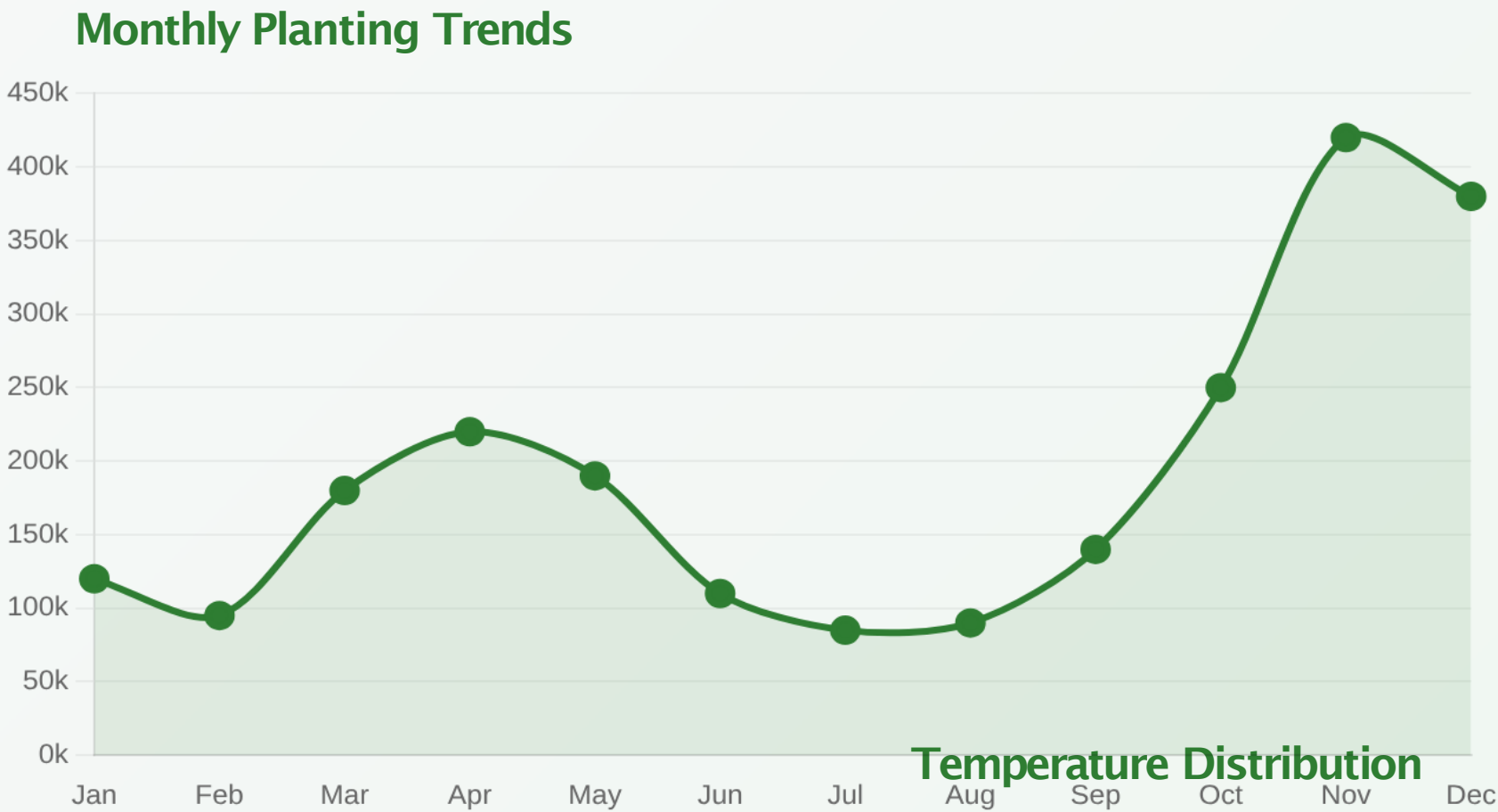
```
df_gps = df_gps[df_gps['seedlingsplanted'] <= 500000]
```

Feature Engineering

Extract month from planting date for seasonal analysis

```
df_gps['month'] = df_gps['plantingdate'].dt.month
```

Exploratory Data Analysis - Temporal Trends



Key Temporal Insights

Coordinated National Efforts

Peaks in **Nov-Dec 2024** suggest coordinated national planting initiatives

Seasonal Alignment

Planting activity increases during **rainy seasons** (OND: Oct-Dec)

Scale Variability

Significant variation in monthly planting volumes, indicating campaign-driven efforts

Environmental Conditions

Average Temperature **24.5°C**

Temperature Range **15°C - 35°C**

Optimal Range **20°C - 28°C**


Planting Success **Higher in moderate temps**


Temperature Impact


Most planting occurs in optimal temperature range, supporting seedling survival

MODEL - Spatial Clustering

Cluster Statistics

 Total Clusters Found **1,642**

 Cluster Radius **1km**

 Minimum Samples **5**

6

Major Hotspots

1,636

Minor Clusters

DBSCAN Methodology

- 1 UTM Coordinate Conversion**
Convert lat/lon to UTM for accurate distance calculations in meters

```
utm_coords = latlon_to_utm(df_gps)
```
- 2 DBSCAN Parameters**
1km radius (eps=1000m) with minimum 5 samples per cluster

```
DBSCAN(eps=1000, min_samples=5, metric='euclidean')
```
- 3 Cluster Assignment**
Assign each planting event to a cluster or mark as outlier (-1)

```
df_gps['cluster'] = cluster.labels_
```

CONCLUSIONS

This analysis reveals that Kenya’s tree planting is concentrated in urban centers, with significant gaps in arid northern counties like Turkana, Wajir, and Mandera. While 78% of planting aligns with rainy seasons—boosting survival—22% still occurs in dry periods, risking seedling loss. Crucially, measuring equity by trees per 10,000 people shows northern counties outperform cities like Nairobi, which ranks only 10th in per-capita impact. The findings call for targeted mobile nurseries, seasonal planting enforcement, and a shift from volume to equity-based metrics. Data-driven strategies can ensure fair, effective, and sustainable reforestation across all regions.

Cluster-Level Performance

Cluster	Rainy Season %	Performance
0	69.6%	Poor
1	66.7%	Poor
2	71.7%	Good
3	45.3%	Poor
4	50.0%	Poor
1637	100.0%	Good
1638	100.0%	Good
1639	100.0%	Good
1640	60.0%	Poor
1641	100.0%	Good

RECOMMENDATION & NEXT STEPS

💡 Key Recommendations

1 Target Under-Planted Counties
Deploy mobile nurseries to northern regions (Mandera, Wajir, Turkana) to address spatial gaps and improve equity

2 Optimize Planting Timing
Schedule all planting activities during MAM (Mar-May) and OND (Oct-Dec) rainy seasons to maximize survival rates

3 Report Per Capita Impact
Shift from volume-based to equity-based metrics (trees per 10k people) for fair assessment across counties

4 Implement Monitoring Dashboard
Create real-time dashboard for monthly tracking of hotspots, gaps, and seasonal alignment performance

★ Transformative Impact



Targeted Action

Data-driven resource allocation to under-served regions



Higher Survival

Improved seedling survival through seasonal optimization



Fair Distribution

Equitable reforestation across all counties



Stakeholder Empowerment

Better decision support for all partners

Project Outcomes:

- ✓ Identified 1,642 planting clusters
- ✓ 78% seasonal alignment achieved
- ✓ Equity framework established
- ✓ Actionable insights for stakeholders

"Reforestation isn't just about planting trees — it's about planting them where they matter most."

THANK YOU