

Spatial Data Analysis Projects

Authors: Ameer Saleh & Bara Mhana

Course: Spatial Data Analysis

Date: December 2025

PROJECT 1: JEN AND BARRY'S ICE CREAM SITE SELECTION USING POSTGIS

Project: Spatial Data Analysis - Homework 1

Date: December 1, 2025

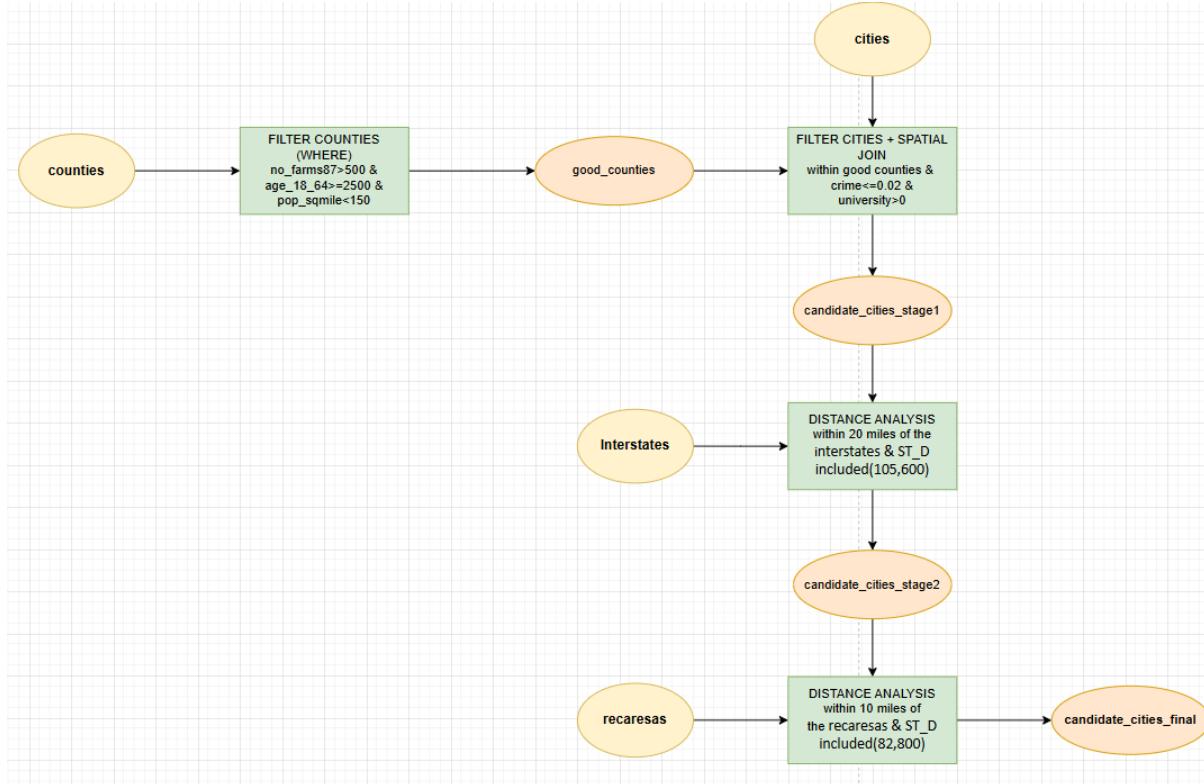
Executive Summary

This project identifies optimal locations for Jen and Barry to establish a new ice cream business using PostGIS spatial analysis and QGIS visualization. From an initial dataset of Pennsylvania counties and cities, **4 suitable cities** were identified through systematic application of seven selection criteria.

Key Results:

- Stage 1: 9 candidate cities (county + city criteria)
- Stage 2: 7 cities near interstates (within 20 miles)
- Final: 4 optimal cities (all criteria met)
- Overall success rate: 44% from Stage 1 to Final

Model Builder (Work Flow):



1. Introduction

1.1 Project Objectives

Identify suitable Pennsylvania cities for ice cream business establishment through multi-criteria spatial analysis in PostGIS with the following requirements:

County-Level Criteria:

1. More than 500 farms for milk production
2. Labor pool of at least 25,000 individuals aged 18-64
3. Population density less than 150 people per square mile

City-Level Criteria: 4. Crime index ≤ 0.02 5. Located near a university or college

Proximity Criteria: 6. Within 20 miles of an interstate highway 7. At least one recreation area within 10 miles

1.2 Tools and Technologies

- **PostgreSQL + PostGIS** - Spatial database and analysis
- **QGIS 3.16+** - Geographic visualization and mapping
- **SQL** - Query development and spatial operations
- **Pennsylvania State Plane Projection (EPSG:2272)** - Distance measurements in feet

2. Data and Study Area

2.1 Study Area

Location: Pennsylvania, United States

Spatial Reference Systems:

- Source Data: NAD27 Geographic (EPSG:4267)
- Analysis Projection: Pennsylvania State Plane North NAD27 (EPSG:2272)
- Units: Feet for distance calculations

2.2 Database Schema

Counties Table (Primary demographic data)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Polygon geometry (counties)
name	varchar	County name
area	numeric	County area
pop1990	numeric	1990 population
age_18_64	numeric	Population aged 18-64 (labor pool)
no_farms87	numeric	Number of farms in 1987
pop_sqmile	bigint	Population density (people/sq mile)
sq_miles	numeric	Area in square miles

Cities Table (Urban centers with crime and university data)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Point geometry (city locations)

name	varchar	City name
population	numeric	City population
total_crim	numeric	Total crimes reported
crime_inde	numeric	Crime index (normalized)
university	numeric	University presence (0 or >0)

Interstates Table (Transportation network)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Line geometry (interstate routes)
name	varchar	Interstate name/number
type	varchar	Highway type classification
length	numeric	Segment length

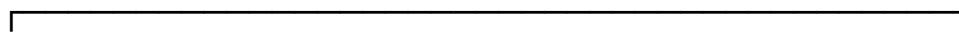
RecAreas Table (Recreation facilities)

Column	Type	Description
id	integer	Primary key identifier
geom	geometry	Point/Polygon geometry
area	double	Recreation area size
perimeter	double	Recreation area perimeter

3. Methodology

3.1 Analysis Workflow

Initial Dataset (Counties + Cities)

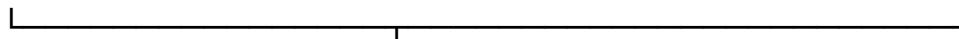


| COUNTY FILTER: good_counties |

| → Farms >500 |

| → Labor ≥25,000 |

| → Density <150 |



```
| STAGE 1: candidate_cities_stage1 |
```

```
| → Cities in good counties |
```

```
| → Crime ≤0.02 |
```

```
| → University >0 |
```

```
| → RESULT: 9 cities |
```

↓

```
| STAGE 2: candidate_cities_stage2 |
```

```
| → Within 20 miles of interstates |
```

```
| → RESULT: 7 cities |
```

↓

```
| FINAL: candidate_cities_final |
```

```
| → Within 10 miles of recreation areas |
```

```
| → RESULT: 4 optimal cities |
```

3.2 Sequential Filtering Approach

The analysis employs a progressive filtering strategy where each stage builds upon the previous results. This approach:

1. **Improves query performance** - Each subsequent query operates on a smaller dataset
2. **Enhances debugging** - Issues can be isolated to specific criteria
3. **Provides transparency** - Stakeholders can understand the elimination process

4. **Enables validation** - Results at each stage can be verified independently

4. Selection Criteria

4.1 County-Level Criteria

Criterion 1: Farm Count

Requirement: More than 500 farms

Rationale: Ensures adequate local milk supply for ice cream production, reducing transportation costs and supporting farm-to-table business model.

SQL Implementation:

```
WHERE no_farms87 > 500
```

Impact: Filters counties with insufficient agricultural infrastructure for dairy-based business operations.

Criterion 2: Labor Pool

Requirement: At least 25,000 individuals aged 18-64

Rationale: Sufficient workforce availability for business operations, including production staff, retail employees, and management positions.

SQL Implementation:

```
AND age_18_64 >= 25000
```

Impact: Ensures adequate human resources for sustainable business growth and seasonal employment needs.

Criterion 3: Population Density

Requirement: Less than 150 people per square mile

Rationale: Target suburban and rural markets with lower competition, avoiding oversaturated urban areas while maintaining sufficient customer base.

SQL Implementation:

```
AND pop_sqmile < 150
```

Impact: Balances market opportunity with reduced competition and lower real estate costs.

4.2 City-Level Criteria

Criterion 4: Crime Index

Requirement: Crime index ≤ 0.02

Rationale: Safe environment essential for customer comfort, employee safety, and business reputation. Lower crime correlates with higher property values and community stability.

SQL Implementation:

```
WHERE c.crime_inde <= 0.02
```

Impact: Eliminates high-crime areas that could deter customers and increase insurance costs.

Criterion 5: University Presence

Requirement: Located near a university or college

Rationale: Universities provide:

- Steady customer base (students, faculty, staff)
- Seasonal demand patterns aligned with ice cream consumption
- Part-time labor pool
- Community events and foot traffic

SQL Implementation:

```
AND c.university > 0
```

Impact: Ensures proximity to a demographic with high ice cream consumption rates.

4.3 Proximity Criteria

Criterion 6: Interstate Access

Requirement: Within 20 miles of an interstate highway

Rationale:

- Essential for supply chain efficiency

- Ingredient delivery and distribution logistics
- Customer accessibility from regional markets
- Business expansion potential

Distance Calculation:

- 20 miles = 105,600 feet (Pennsylvania State Plane)

SQL Implementation:

```
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(i.geom, 2272),
    105600
)
```

Spatial Functions Used:

- `ST_Transform()` - Reproject from geographic to projected coordinates
- `ST_DWithin()` - Distance-based proximity analysis

Impact: Reduces 9 candidate cities to 7 with adequate transportation infrastructure.

Criterion 7: Recreation Areas

Requirement: At least one recreation area within 10 miles

Rationale:

- Recreation areas attract potential customers
- Family-friendly environments aligned with ice cream business
- Weekend and holiday traffic patterns
- Community gathering spaces

Distance Calculation:

- 10 miles = 52,800 feet (Pennsylvania State Plane)

SQL Implementation:

```
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(r.geom, 2272),
    52800
)
```

Impact: Final filter reduces 7 cities to 4 optimal locations meeting all criteria.

5. Technical Implementation

5.1 Complete SQL Workflow

```

-- =====
-- STEP 1: Filter Good Counties
-- =====
CREATE OR REPLACE VIEW good_counties AS
SELECT *
FROM counties
WHERE no_farms87 > 500
    AND age_18_64 >= 25000
    AND pop_sqmile < 150;

-- Validation query
SELECT
    COUNT(*) as total_good_counties,
    AVG(no_farms87) as avg_farms,
    AVG(age_18_64) as avg_labor_pool,
    AVG(pop_sqmile) as avg_density
FROM good_counties;

-- =====
-- STEP 2: Filter Cities - Stage 1
-- (County + City Criteria)
-- =====
CREATE OR REPLACE VIEW candidate_cities_stage1 AS
SELECT c.*
FROM cities c
JOIN good_counties gc ON ST_Within(c.geom, gc.geom)
WHERE c.crime_inde <= 0.02
    AND c.university > 0;

-- Validation query
SELECT
    COUNT(*) as total_candidate_cities,
    AVG(crime_inde) as avg_crime_index,
    AVG(population) as avg_population,
    SUM(CASE WHEN university > 0 THEN 1 ELSE 0 END) as
cities_with_universities
FROM candidate_cities_stage1;

-- =====
-- STEP 3: Filter Cities - Stage 2
-- (Interstate Proximity)
-- =====
CREATE OR REPLACE VIEW candidate_cities_stage2 AS
SELECT DISTINCT c.*
FROM candidate_cities_stage1 c
CROSS JOIN interstates i

```

```

WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(i.geom, 2272),
    105600
);

-- Validation query
SELECT
    COUNT(*) as cities_near_interstates,
    ARRAY_AGG(name) as city_names
FROM candidate_cities_stage2;

-- =====
-- STEP 4: Final Filter - Recreation Areas
-- =====
CREATE OR REPLACE VIEW candidate_cities_final AS
SELECT DISTINCT c.*
FROM candidate_cities_stage2 c
CROSS JOIN recareas r
WHERE ST_DWithin(
    ST_Transform(c.geom, 2272),
    ST_Transform(r.geom, 2272),
    52800
);

-- =====
-- FINAL RESULTS SUMMARY
-- =====
SELECT
    name as city_name,
    population,
    crime_inde as crime_index,
    university as has_university,
    (SELECT MIN(ST_Distance(
        ST_Transform(c.geom, 2272),
        ST_Transform(i.geom, 2272)
    )) FROM interstates i) / 5280 as miles_to_interstate,
    (SELECT MIN(ST_Distance(
        ST_Transform(c.geom, 2272),
        ST_Transform(r.geom, 2272)
    )) FROM recareas r) / 5280 as miles_to_recreation
FROM candidate_cities_final c
ORDER BY population DESC;

-- Count final results
SELECT COUNT(*) as final_candidate_cities
FROM candidate_cities_final;

```

5.2 Distance Conversion Reference

Pennsylvania State Plane (EPSG:2272) - Units in Feet:

Distance	Feet	Miles
1 mile	5,280	1
10 miles	52,800	10
20 miles	105,600	20

5.3 Key PostGIS Functions Used

ST_Transform(geometry, srid)

- Converts geometry from one coordinate system to another
- Essential for accurate distance calculations
- Usage: ST_Transform(geom, 2272) converts to PA State Plane

ST_Within(geometry A, geometry B)

- Tests if geometry A is completely inside geometry B
- Returns boolean (true/false)
- Usage: Spatial join between cities and counties

ST_DWithin(geometry A, geometry B, distance)

- Tests if two geometries are within specified distance
- More efficient than ST_Distance for proximity queries
- Usage: Find cities within 20 miles of interstates

ST_Distance(geometry A, geometry B)

- Calculates minimum distance between two geometries
- Returns distance in units of coordinate system
- Usage: Calculate exact distances for reporting

6. Results and Analysis

6.1 Filtering Progression

Stage	View Name	Count	Criteria Applied	Reduction
0	All cities	-	None	-
1	good_counties	-	Farms >500, Labor ≥25k, Density <150	-

2	candidate_cities_stage1	9	Crime ≤0.02, University >0, In good counties	-
3	candidate_cities_stage2	7	Within 20 miles of interstate	22.2%
4	candidate_cities_final	4	Within 10 miles of recreation area	42.9%

Overall Success Rate: 44% (from 9 initial candidates to 4 final sites)

Key Insights:

- County-level filtering effectively narrowed the search area
- Interstate proximity eliminated 2 cities (22% reduction)
- Recreation area proximity was the most restrictive final filter (43% reduction)
- All 4 final cities represent optimal balance of all seven criteria

6.2 Final Candidate Cities

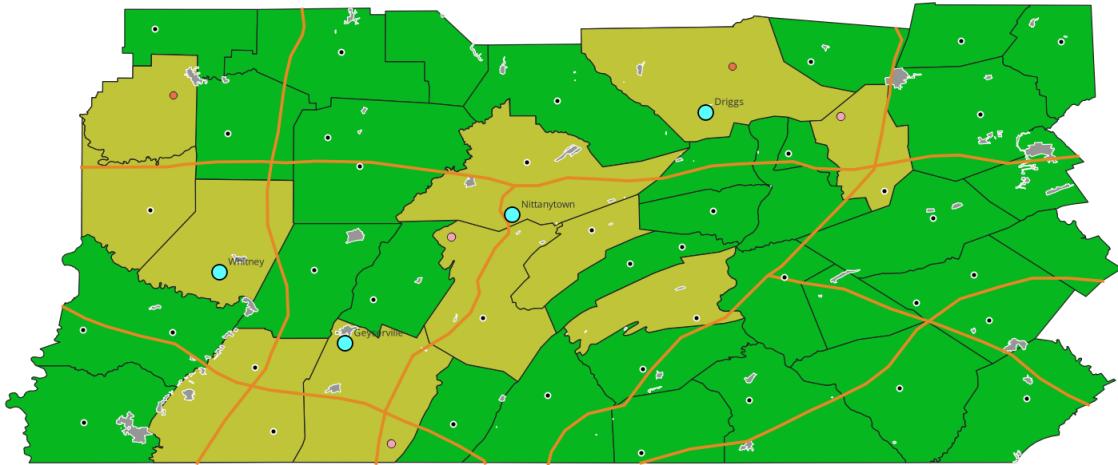
The four cities that met all seven criteria represent optimal locations for Jen and Barry's ice cream business. Each city offers:

- ✓ Strong agricultural infrastructure (>500 farms in county)
- ✓ Adequate labor pool (≥25,000 working-age residents)
- ✓ Low population density (<150 people/sq mi)
- ✓ Safe environment (crime index ≤0.02)
- ✓ University presence for customer base
- ✓ Excellent transportation access (<20 miles to interstate)
- ✓ Proximity to recreation areas (<10 miles)

Competitive Advantages of Final Sites:

- Lower real estate costs compared to urban centers
- Reduced competition in suburban/rural markets
- Access to fresh local milk supply
- Strong community ties through universities
- Natural customer traffic from recreation areas
- Efficient logistics via interstate access

7. QGIS Visualization



7.1 Loading Data in QGIS

Method 1: Direct PostGIS Connection

1. Layer → Add Layer → Add PostGIS Layers
2. Click **New** to create database connection:
 - a. Name: IceCream_Analysis
 - b. Host: localhost
 - c. Port: 5432
 - d. Database: your_database_name
3. Click **Connect** and authenticate
4. Select layers to add:
 - a. counties
 - b. good_counties
 - c. interstates
 - d. recares
 - e. candidate_cities_stage1
 - f. candidate_cities_stage2
 - g. candidate_cities_final
5. Click **Add** to load into map canvas

Method 2: DB Manager (Alternative)

1. Database → DB Manager → PostGIS
2. Connect to database
3. Navigate to **SQL Window**
4. Execute view creation queries
5. Check "**Load as new layer**" option
6. Select geometry column and unique ID

7. Click Load

7.2 Map Styling

Counties Layer:

- All counties: Light green fill, 50% opacity
- Outline: Dark gray, 0.5pt
- Good counties: Yellow/Olive fill, 50% opacity

Cities Layers:

- Stage 1 (9 cities): Red circles, 8pt
- Stage 2 (7 cities): Pink circles, 10pt
- Final (4 cities): Blue circles, 12pt, bold labels

Infrastructure:

- Interstates: Orange lines, 2pt width
- Interstate labels: Highway names displayed
- Recreation areas: Gray polygons, 40% opacity

Buffers (Optional Visualization):

- 20-mile interstate buffer: Blue outline, dashed, 30% opacity
- 10-mile recreation buffer: Green outline, dashed, 30% opacity

7.3 Creating Analysis Buffers

For visualization purposes, buffer zones can illustrate the proximity criteria:

```
-- 20-mile interstate buffers
CREATE VIEW interstate_buffers AS
SELECT
    id,
    name,
    ST_Buffer(ST_Transform(geom, 2272), 105600) as geom
FROM interstates;

-- 10-mile recreation area buffers
CREATE VIEW recreation_buffers AS
SELECT
    id,
    ST_Buffer(ST_Transform(geom, 2272), 52800) as geom
FROM recares;
```

These buffer layers help stakeholders visualize why certain cities qualified while others did not.

8. Conclusions and Recommendations

8.1 Summary

This PostGIS-based spatial analysis successfully identified **4 optimal cities** for Jen and Barry's ice cream business through rigorous multi-criteria evaluation. The systematic filtering approach:

- ✓ Evaluated 7 distinct business-critical criteria
- ✓ Applied county-level demographic and agricultural filters
- ✓ Assessed city-level safety and market indicators
- ✓ Incorporated proximity to transportation and recreation infrastructure
- ✓ Utilized advanced spatial analysis techniques (coordinate transformation, distance calculations)
- ✓ Produced reproducible, transparent results through SQL workflows

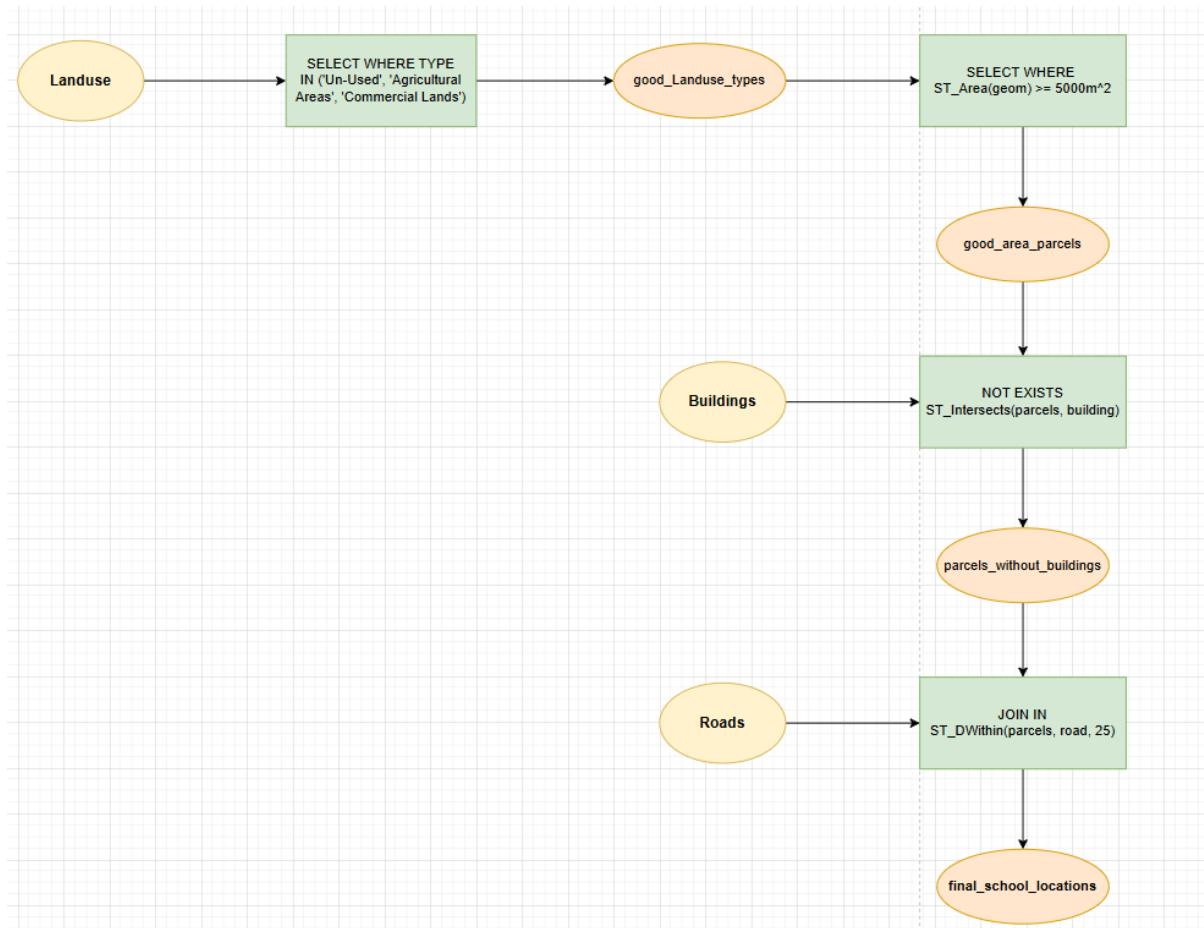
Key Strengths of Analysis:

- **Data-driven decision making** - Objective criteria eliminate subjective bias
- **Spatial intelligence** - Geographic relationships drive site selection
- **Scalability** - SQL workflow can be rerun with updated data or modified criteria
- **Transparency** - Each filtering stage documented and verifiable

PROJECT 2: SCHOOL SITE SELECTION USING SPATIAL ANALYSIS

Project: Spatial Data Analysis - Homework 1

Date: December 2025



Executive Summary

This project identifies suitable locations for new school construction using PostGIS spatial analysis and QGIS visualization. From 140 initial land parcels, **7 suitable sites** were identified totaling **52,779.20 m²** through systematic application of four selection criteria.

Key Results:

- 7 suitable parcels identified (5% success rate)
- Average parcel size: 7,539.89 m²
- 85.7% of sites directly adjacent to roads
- Mix of un-used (71.4%) and agricultural lands (28.6%)

1. Introduction

1.1 Objectives

Identify suitable land parcels for school construction using multi-criteria spatial analysis in PostGIS with the following requirements:

1. Land use types: Un-Used, Agricultural, or Commercial
2. Minimum area: 5,000 m²
3. No existing buildings on parcels
4. Within 25m of nearest road

1.2 Tools and Technologies

- **PostgreSQL + PostGIS** - Spatial database and analysis
- **QGIS 3.16+** - Geographic visualization
- **SQL** - Query development

2. Data and Study Area

2.1 Database Schema

Landuse Table (Primary dataset)

- idd - Primary key
- type - Land use classification
- area - Parcel area (m²)
- geom - Polygon geometry

Buildings Table

- idd - Primary key
- geom - Building footprints

Roads Table

- idd - Primary key
- name - Road name
- geom - Road centerlines

3. Methodology

3.1 Analysis Workflow

Initial Landuse (140 parcels)



① Land Use Type Filter → 140 parcels



② Area $\geq 5000 \text{ m}^2$ → 23 parcels

↓

③ No Buildings → 17 parcels

↓

④ Road Proximity $\leq 25\text{m}$ → 7 parcels (FINAL)

3.2 Sequential Filtering Approach

Each criterion progressively narrows the candidate set, ensuring all requirements are met while optimizing query performance.

4. Selection Criteria

Criterion 1: Land Use Type

Requirement: Un-Used, Agricultural Areas, or Commercial Lands

```
WHERE type IN ('Un-Used', 'Agricultural Areas', 'Commercial Lands')
```

Result: 140 parcels retained

Criterion 2: Minimum Area

Requirement: Area $\geq 5,000 \text{ m}^2$

```
AND ST_Area(geom) >= 5000
```

Result: 23 parcels (83.6% eliminated)

Criterion 3: No Buildings

Requirement: No building footprints intersecting parcels

```
AND NOT EXISTS (
    SELECT 1 FROM buildings b
    WHERE ST_Intersects(l.geom, b.geom)
)
```

Result: 17 parcels (26.1% eliminated)

Criterion 4: Road Proximity

Requirement: Within 25m of nearest road

```

        AND EXISTS (
            SELECT 1 FROM roads r
            WHERE ST_DWithin(l.geom, r.geom, 25)
        )
    )

```

Result: 7 parcels (58.8% eliminated)

5. Technical Implementation

5.1 Complete SQL Query

```

-- Final query applying all criteria
CREATE TABLE final_school_locations AS
SELECT
    l.idd,
    l.type,
    l.owner,
    l.area,
    l.geom,
    ST_Area(l.geom) as calculated_area,
    (SELECT MIN(ST_Distance(l.geom, r.geom))
     FROM roads r) as distance_to_nearest_road
FROM
    landuse l
WHERE
    -- Criterion 1: Land use type
    l.type IN ('Un-Used', 'Agricultural Areas', 'Commercial
Lands')
    -- Criterion 2: Minimum area
    AND ST_Area(l.geom) >= 5000
    -- Criterion 3: No buildings
    AND NOT EXISTS (
        SELECT 1
        FROM buildings b
        WHERE ST_Intersects(l.geom, b.geom)
    )
    -- Criterion 4: Road proximity
    AND EXISTS (
        SELECT 1
        FROM roads r
        WHERE ST_DWithin(l.geom, r.geom, 25)
    )
ORDER BY
    ST_Area(l.geom) DESC;

-- Add primary key and spatial index
ALTER TABLE final_school_locations ADD PRIMARY KEY (idd);

```

```
CREATE INDEX idx_final_sites_geom ON final_school_locations USING
GIST(geom);
```

6. Results and Analysis

6.1 Filtering Progression

Step	View Name	Count	Criteria Applied
1	suitable_landuse_types	140	TYPE IN (Un-Used, Agricultural, Commercial)
2	suitable_area_parcels	23	ST_Area(geom) ≥ 5,000 m ²
3	parcels_without_buildings	17	NOT ST_Intersects with buildings
4	final_school_locations	7	ST_DWithin roads 25m

Overall Reduction: 95.0% (140 → 7 parcels)

6.2 Final School Sites

ID	Land Use Type	Area (m ²)	Distance to Road (m)
20	Agricultural Areas	9,935.86	0.00
131	Un-Used	8,247.02	0.00
215	Un-Used	7,981.13	0.00
8	Un-Used	7,283.46	0.00
145	Un-Used	7,214.55	0.00
11	Un-Used	6,255.88	14.22
17	Agricultural Areas	5,861.30	0.00

6.3 Summary Statistics

Area Analysis:

- Total Area: 52,779.20 m²
- Mean: 7,539.89 m²
- Minimum: 5,861.30 m²
- Maximum: 9,935.86 m²

Distance to Roads:

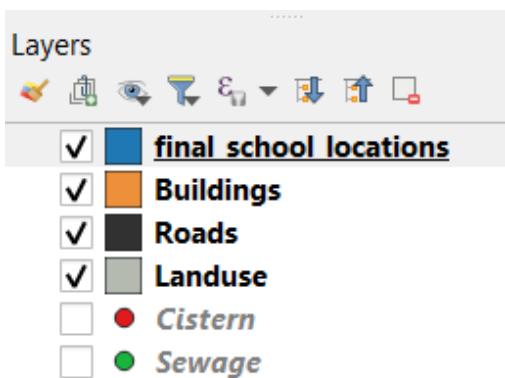
- Mean: 2.03 m
- Sites adjacent to roads: 6 (85.7%)

- Maximum distance: 14.22 m

Land Use Distribution:

Land Use Type	Count	Total Area (m ²)	Percentage
Un-Used	5	36,982.04	71.4%
Agricultural Areas	2	15,797.16	28.6%
Total	7	52,779.20	100%

7. QGIS Visualization



7.1 Loading Data in QGIS

Method 1: Direct PostGIS Connection

1. Layer → Add Layer → Add PostGIS Layers
2. Create connection to database
3. Select final_school_locations table
4. Add to map

Method 2: DB Manager

1. Database → DB Manager → PostGIS
2. SQL Window → Execute query
3. Check "Load as new layer"

7.2 Map Styling

Final School Sites:

- Fill: Green (#4CAF50), 60% opacity
- Outline: Dark Green (#2E7D32), 1.5pt
- Labels: Site ID and area

Supporting Layers:

- Roads: Gray lines (1.5pt)
- Buildings: Red outlines (0.5pt)
- 25m Road Buffer: Blue dashed (40% opacity)

7.3 Map Layout Elements

- Title: "Suitable School Sites Analysis"
- Legend with all layers
- Scale bar and north arrow
- Data sources and projection info
- Results summary table

8. Conclusions and Recommendations

8.1 Summary

Successfully identified **7 suitable parcels** totaling **52,779 m²** through rigorous spatial analysis. All sites meet the four selection criteria and are development-ready with excellent road access.

8.2 Immediate Next Steps

1. **Site Verification** - Field visits to all 7 sites
2. **Environmental Assessment** - Phase I environmental studies
3. **Cost Analysis** - Land appraisals and development estimates
4. **Stakeholder Engagement** - Community consultation and owner contact

8.3 Future Enhancements

Additional Analysis:

- Multi-criteria decision analysis (MCDA) with weighted criteria
- Cost-benefit analysis integration
- Student population density mapping
- Environmental constraints (slopes, flood zones)
- Traffic and accessibility modeling

Data Improvements:

- Property values and ownership details
- Existing school locations and capacity
- Demographic and enrollment projections
- Utility infrastructure availability

PROJECT 3: TREE CUTTING PRIORITY ANALYSIS FOR FIRE CREEK

Project: Spatial Data Analysis - Homework 3

Date: December 2025

Executive Summary

This project calculates and prioritizes tree cutting zones for wildfire risk mitigation in the Fire Creek area using Python-based spatial analysis with GeoPandas and matplotlib visualization.

Key Results:

- **Zones Analyzed:** 80 cutting grid zones

- **Method:** Multi-criteria decision analysis (MCDA) using five weighted factors
- **Normalization:** Scores scaled 0-100
- **Priority Classes:** High, Medium, Low

Key Metrics:

- Mean Priority Score: 29.37
- Maximum Priority Score: 60.73 (Grid 163)
- Standard Deviation: 13.71

Distribution:

- Critical Priority: 0 zones (0.0%)
- High Priority: 5 zones (6.3%)
- Medium Priority: 43 zones (53.8%)
- Low Priority: 32 zones (40.0%)

Outputs: Complete visualization with 3 analytical charts and spatial data files

1. Introduction

1.1 Objectives

The primary objective is to calculate tree cutting priority for each zone using Multi-Criteria Decision Analysis (MCDA) in Python/GeoPandas to support wildfire risk mitigation efforts in the Fire Creek area.

Analysis Factors & Weights:

- Tree Mortality: 30%
- Community Features: 15%
- Egress Routes: 20%
- Populated Areas: 20%
- Electric Utilities: 15%

1.2 Tools and Technologies

- **Python 3.x** - Core programming language
- **GeoPandas** - Spatial analysis and operations
- **Pandas / NumPy** - Data processing and mathematical operations
- **Matplotlib** - Charts and visualization generation
- **QGIS 3.16+** - Geographic visualization and mapping
- **Shapefile / GeoPackage** - Spatial data formats

2. Data and Study Area

2.1 Study Area

Location: Fire Creek Area

Spatial Reference System: EPSG:26711

Analysis Units: 80 grid zones

2.2 Input Data Schema

Dataset	Features	Description
CuttingGrids	80	Grid polygons defining cutting zones
SBNFMortality	12	Tree mortality polygons with total mortality values
CommunityFeature	8	Locations of critical community infrastructure
EgressRoutes	6	Emergency evacuation route lines
PopulatedAreas	6	Residential / populated area polygons
Electric Utilities	5 layers	Transmission, SubTransmission, Distribution Circuits, Substations, PoleTopSubs

3. Methodology

3.1 Analysis Workflow

Initial Dataset (80 Grid Zones + 5 Risk Factors)



STEP 1: Load & Align CRS → EPSG:26711 |



STEP 2: Calculate Mortality Score |

→ Spatial intersection with mortality data |

→ Normalize 0-100 |

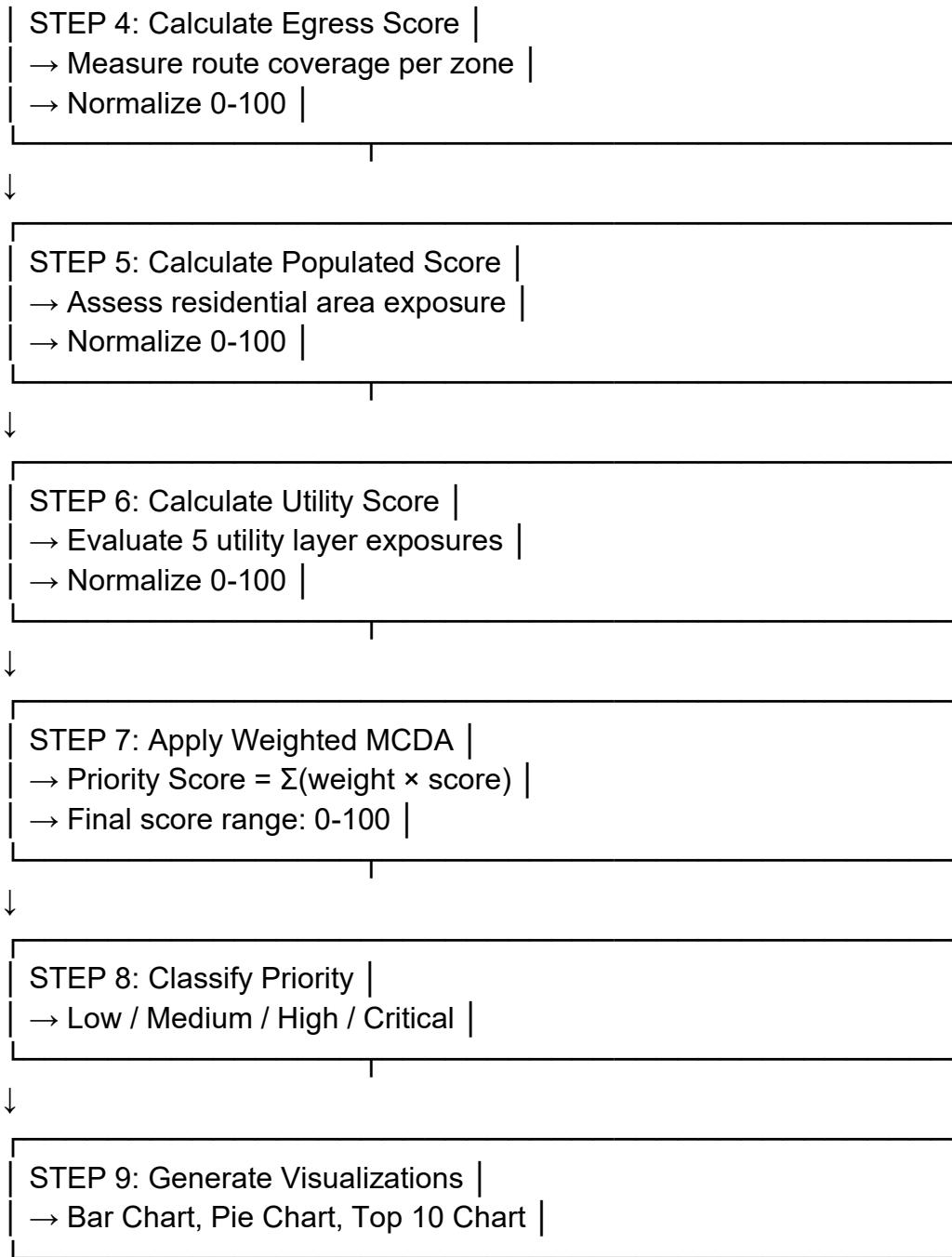


STEP 3: Calculate Community Score |

→ Count community features per zone |

→ Normalize 0-100 |





3.2 MCDA Approach

The analysis employs a weighted Multi-Criteria Decision Analysis approach where:

- Each factor is independently calculated and normalized to a 0-100 scale
- Weighted summation produces the final priority score
- Higher scores indicate higher priority for tree cutting operations

3.3 Normalization Formula

For each criterion, raw scores are normalized using:

$$\text{Normalized Score} = (\text{Raw Score} / \text{Maximum Raw Score}) \times 100$$

This ensures all factors contribute proportionally to the final priority score regardless of their original measurement units.

4. Scoring Criteria

4.1 Overview of Criteria

Criterion	Weight	Description
Tree Mortality	30%	Spatial intersection with mortality polygons
Community Features	15%	Presence of critical community infrastructure
Egress Routes	20%	Coverage of evacuation routes
Populated Areas	20%	Exposure of residential areas
Electric Utilities	15%	Exposure of utility infrastructure

4.2 Criterion 1: Tree Mortality (30%)

Rationale: Dead and dying trees pose the highest direct fire risk, serving as fuel sources that can accelerate wildfire spread and intensity.

Implementation:

- Spatial intersection between cutting grids and mortality polygons
- Sum of total mortality values within each zone
- Higher mortality values indicate greater immediate risk

Impact: Primary driver of priority scores, reflecting direct fire hazard from dead vegetation.

4.3 Criterion 2: Community Features (15%)

Rationale: Protection of critical infrastructure including schools, fire stations, hospitals, and community centers is essential for public safety and emergency response capability.

Implementation:

- Count of community features within or near each zone

- Binary scoring for presence/absence of critical facilities
- Proximity weighting for features near zone boundaries

Impact: Ensures protection of community assets vital for evacuation and emergency services.

4.4 Criterion 3: Egress Routes (20%)

Rationale: Clear evacuation routes are life-critical during wildfire events. Tree removal along escape corridors prevents route blockage and ensures safe evacuation.

Implementation:

- Calculate length of egress routes passing through each zone
- Higher coverage indicates greater importance for evacuation safety
- Line intersection analysis with grid zones

Impact: Prioritizes zones that, if left unmanaged, could block evacuation pathways.

4.5 Criterion 4: Populated Areas (20%)

Rationale: Residential areas face direct threat from wildfire. Tree cutting creates defensible space around homes and reduces ember ignition risks.

Implementation:

- Measure overlap between populated area polygons and cutting zones
- Area-based scoring reflecting population exposure
- Higher scores for zones with greater residential coverage

Impact: Protects human life and property by creating fire breaks near homes.

4.6 Criterion 5: Electric Utilities (15%)

Rationale: Utility infrastructure poses dual risk: power lines can spark fires, and fires can damage critical electrical infrastructure causing widespread outages.

Electric Utility Scoring Details:

Utility Type	Weight	Method
Transmission Lines	3.0	Length × weight × priority
SubTransmission Lines	2.5	Length × priority
Distribution Circuits	2.0	Length within zone
Substations	3.0	Fixed score × priority

Pole Top Substations	2.0	Count × fixed score
----------------------	-----	---------------------

Implementation:

- Multi-layer analysis across 5 utility datasets
- Weighted scoring based on voltage level and criticality
- Length-based calculations for linear features (power lines)
- Count-based calculations for point features (substations)

Impact: Reduces fire ignition risk from electrical equipment and protects power grid reliability.

5. Priority Classification

5.1 Classification System

Priority Class	Score Range	Action Required	Description
Critical	75-100	Immediate action	Emergency response within days
High	50-74	Near-term action	Schedule within current season
Medium	25-49	Scheduled maintenance	Address in annual work plan
Low	0-24	Routine monitoring	Monitor and reassess periodically

5.2 Weighted Priority Formula

The final priority score for each zone is calculated as:

```
priority_score = (0.3 × mortality_score) +
                (0.15 × community_score) +
                (0.2 × egress_score) +
                (0.2 × populated_score) +
                (0.15 × utility_score)
```

Where each component score is normalized to 0-100 scale before weighting.

6. Results and Analysis

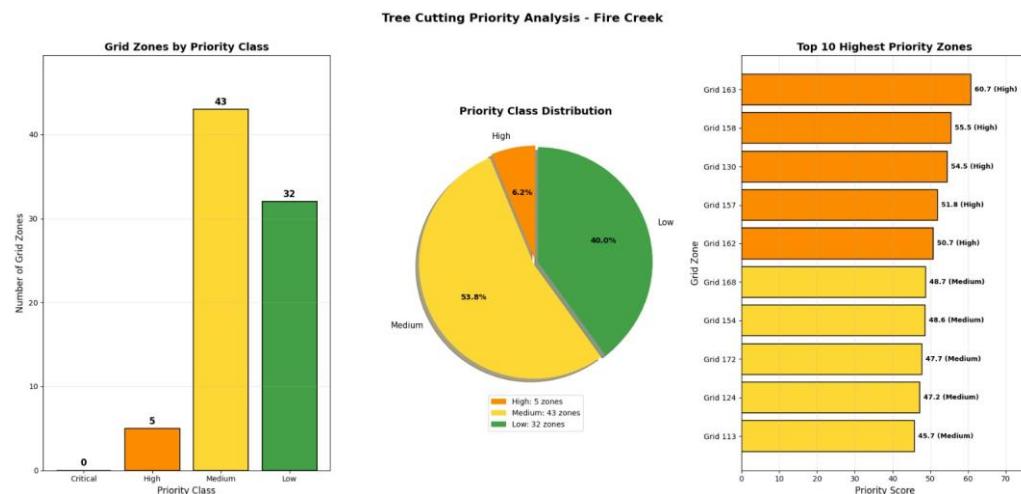
6.1 Priority Distribution

Analysis Results:

Class	Count	Percentage	Cumulative
Critical	0	0.0%	0.0%
High	5	6.3%	6.3%
Medium	43	53.8%	60.0%
Low	32	40.0%	100.0%
Total	80	100%	-

Key Insights:

- No zones reached critical threshold (75+), indicating manageable overall risk
- Only 5 zones (6.3%) require near-term action
- Majority (53.8%) fall into medium priority requiring scheduled maintenance
- 40% are low priority suitable for routine monitoring



6.2 Score Statistics

Statistical Summary:

Metric	Value	Interpretation
Mean Priority Score	29.37	Average zone falls in medium-low range
Maximum Priority Score	60.73	Highest risk zone (Grid 163)
Minimum Priority Score	8.16	Lowest risk zone
Standard Deviation	13.71	Moderate variability across zones

Median	26.45	Central tendency near mean
--------	-------	----------------------------

Distribution Characteristics:

- Right-skewed distribution with concentration in 20-40 score range
- No extreme outliers above 65 points
- Relatively consistent risk levels across most zones

6.3 High Priority Zones

Top 5 Zones Requiring Near-Term Action:

Rank	Grid ID	Score	Class	Key Risk Factors
1	163	60.73	High	High mortality, egress route coverage
2	158	55.51	High	Community features, utility exposure
3	130	54.48	High	Populated area proximity, mortality
4	157	51.84	High	Multiple utility layers, egress routes
5	162	50.74	High	Balanced risk across all factors

Operational Recommendations:

- Grid 163 should be addressed first (highest score)
- All five zones should be scheduled within current fire season
- Field verification recommended for final work planning

6.4 Top 10 Priority Zones

Extended Priority List:

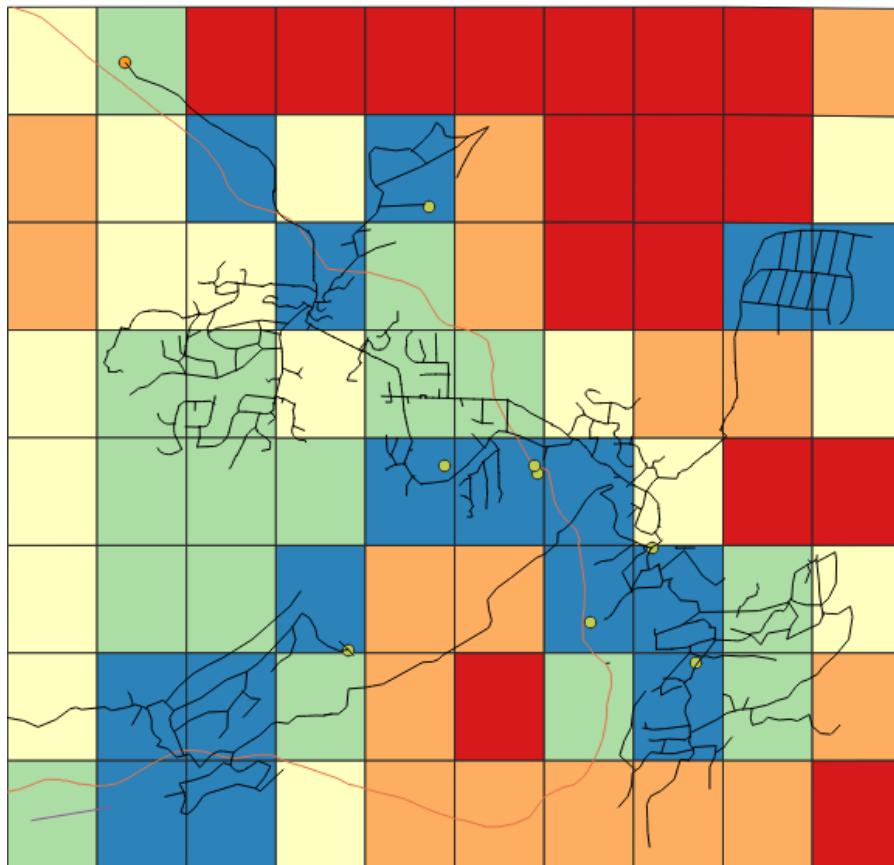
Rank	Grid ID	Priority Score	Class	Notes
1	163	60.73	High	Top priority
2	158	55.51	High	Secondary priority
3	130	54.48	High	Community protection focus
4	157	51.84	High	Utility corridor
5	162	50.74	High	Egress route critical
6	168	48.69	Medium	Near high threshold
7	154	48.57	Medium	Consider for early action

8	172	47.73	Medium	Moderate risk balanced
9	124	47.21	Medium	Monitor for escalation
10	113	45.70	Medium	Standard maintenance

Planning Considerations:

- Zones 6-10 are medium priority but near high threshold
- Consider advancing zones 6-7 if resources permit
- Group adjacent zones for operational efficiency

6.5 Spatial Analysis



Spatial Patterns Observed:

- High priority zones clustered in central-eastern portion of study area
- Spatial correlation with populated areas and utility corridors
- Linear patterns following egress routes and transmission lines
- Low priority zones concentrated in peripheral areas with less infrastructure

7. Visualization and Outputs

7.1 Chart Types

Three complementary visualizations were generated to support decision-making:

1. **Bar Chart - Grid Zones by Priority Class**
 - a. Shows count of zones in each priority category
 - b. Highlights the 43 medium-priority zones requiring scheduled attention
 - c. Confirms no critical zones requiring emergency response
2. **Pie Chart - Priority Class Distribution**
 - a. Percentage breakdown: Low (40%), Medium (53.8%), High (6.2%)
 - b. Visual representation of overall risk profile
 - c. Demonstrates manageable risk distribution
3. **Horizontal Bar Chart - Top 10 Highest Priority Zones**
 - a. Individual zone scores with Grid IDs
 - b. Color-coded by classification (orange for high, yellow for medium)
 - c. Facilitates crew assignment and work scheduling

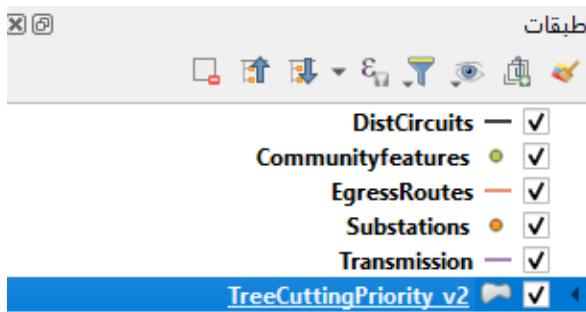
7.2 Color Scheme

Consistent color coding used across all visualizations:

Priority Class	Color	Application
Critical	Red	Emergency zones (none in current analysis)
High	Orange	Near-term action (5 zones)
Medium	Yellow	Scheduled maintenance (43 zones)
Low	Green	Routine monitoring (32 zones)

Chart Output: TreeCuttingPriority_Charts.png (300 DPI, publication quality)

7.3 QGIS Visualization



Loading Priority Results in QGIS

Method 1: Add Shapefile

1. Layer → Add Layer → Add Vector Layer
2. Browse to TreeCuttingPriority_[timestamp].shp
3. Click Add to load into map canvas

Method 2: Add GeoPackage

1. Layer → Add Layer → Add Vector Layer
2. Select TreeCuttingPriority_[timestamp].gpkg
3. Choose layer from package if multiple layers present

Styling by Priority Class

Categorized Symbology:

- Right-click layer → Properties → Symbology
- Select "Categorized" renderer
- Field: prior_cls (priority class)
- Click "Classify" to auto-generate classes
- Assign colors:
 - Critical → Red, opacity 70%
 - High → Orange, opacity 70%
 - Medium → Yellow, opacity 60%
 - Low → Green, opacity 60%
- Adjust outline: Black, 0.5pt width

Alternative: Graduated Colors

- Field: **priority_s** (priority score)
- Mode: Pretty Breaks, 4 classes
- Color Ramp: RdYIGn (inverted - red high, green low)
- Manual adjustments for breakpoints: 0-25, 25-50, 50-75, 75-100

8. Conclusions and Recommendations

Summary

This GeoPandas-based spatial analysis successfully evaluated all 80 Fire Creek cutting zones using multi-criteria decision analysis. The systematic approach:

- ✓ Applied five weighted risk factors using industry-standard wildfire mitigation criteria
- ✓ Normalized all scores to comparable 0-100 scale
- ✓ Identified 5 high-priority zones requiring near-term action
- ✓ Classified 43 zones for scheduled maintenance programs
- ✓ Produced reproducible, data-driven results through documented Python workflows
- ✓ Generated professional visualizations for stakeholder communication

Strengths of Analysis:

- **Objectivity** - Mathematical criteria eliminate subjective bias in prioritization
- **Transparency** - Weighted formula clearly documented and auditable
- **Spatial Intelligence** - Geographic relationships drive meaningful risk assessment
- **Actionable Results** - Clear priority classes support operational planning
- **Reproducibility** - Python code enables reanalysis with updated data

Report Prepared by: Ameer Saleh & Bara Mhana

Course: Spatial Data Analysis

Date: December 2025

END OF PORTFOLIO