Final Project Proposal: Comfort-Optimized Routes to Transit

Enhancing Pedestrian Comfort and Access to Public Transportation in Philadelphia

Student: [Your Name]

Course: MUSA 5500 - Geospatial Data Science in Python

Date: November 6, 2024

Group Size: Solo (2 requirements minimum)

---

1. Project Overview

Problem Statement

As urban temperatures rise due to climate change, pedestrian comfort during daily commutes has become a critical concern for public health and transportation planning. Transit-dependent populations are particularly vulnerable, as they must walk to bus stops and train stations regardless of environmental conditions. Pedestrian comfort is influenced by multiple factors including shade coverage, street width, building heights, surface materials, and urban greenery - yet conventional routing applications optimize only for distance or time.

This project addresses a fundamental question: **How can we develop comfort-aware pedestrian routes to transit stops that balance environmental factors with reasonable travel distances?**

Research Questions

1. How do pedestrian comfort factors (tree canopy, street characteristics, urban form) vary along routes to transit stops across Philadelphia?
2. What is the distance/time trade-off between direct routes and comfort-optimized routes to the nearest transit stops?
3. Which neighborhoods have the poorest comfortable access to transit, and where should the city prioritize environmental improvements?
4. Are there disparities in comfortable transit access across different demographic groups?

Significance

This project provides actionable insights for:

- **Urban planners** prioritizing environmental improvements along high-traffic pedestrian corridors
- **Transit agencies** understanding "last mile" comfort challenges that affect ridership
- **Public health officials** identifying vulnerable populations facing daily environmental stressors

- **Technology developers** building comfort-aware navigation applications
- **Climate adaptation planners** designing interventions for pedestrian-friendly streets

---

2. Project Requirements

Required Elements (5 for A+ Ambition)

1. OSMnx for Street Network Analysis ✅

- Extract walkable street network for Philadelphia
- Calculate pedestrian routes from sample locations to nearest transit stops
- Implement custom routing algorithm with comfort-weighted edges
- Compute walking distance and time for route comparisons

2. Raster Data Analysis ✅

- Process Philadelphia Land Cover Raster (2018) to extract tree canopy and surface types
- Use `rasterio` to read and manipulate raster data
- Use `rasterstats` to sample environmental characteristics along street segments
- Generate comfort scores for each network edge based on multiple factors

3. Multiple Data Sources (3+) ✅

- OpenStreetMap street network (via OSMnx)
- Philadelphia Land Cover Raster - tree canopy and surface types (OpenDataPhilly)
- SEPTA GTFS transit data (stops, routes, service frequency)
- U.S. Census demographic data (income, race, car ownership)
- *Four distinct data sources from different providers*

4. Reasonably Complex Geospatial Analysis ✅

- Multi-factor comfort scoring combining raster and vector data
- Spatial joins between transit stops and street network
- Buffer analysis and zonal statistics along routes
- Network routing with custom multi-criteria weights

- Isochrone generation for accessibility analysis
- Spatial overlay with demographic data for equity analysis

5. Deployed Panel Dashboard ✅ *(Stretch Goal)*

- Interactive web application for route exploration
- User inputs: origin address, preferred comfort vs. distance trade-off
- Dynamic visualization: comfort-optimized routes to 3 nearest transit stops
- Adjustable weights for comfort factors (trees, street width, surfaces)
- Comparison metrics displayed in real-time
- Deployable on free hosting (e.g., Hugging Face Spaces, Railway.app)

---

3. Data Sources

Primary Data (Confirmed Available)

Street Network

- **Source:** OpenStreetMap via OSMnx
- **Coverage:** Complete Philadelphia street network
- **Access:** Free, programmatic access through Python
- **Format:** NetworkX graph object

Tree Canopy Coverage

- **Source:** Philadelphia Land Cover Raster (2018)
- **Provider:** OpenDataPhilly / City of Philadelphia
- **Resolution:** 1-meter resolution GeoTIFF
- **Coverage:** Complete city coverage with 7 land cover classes including tree canopy, grass/shrub, bare earth, water, buildings, roads, and other paved surfaces
- **Access:** Free download from OpenDataPhilly.org
- **URL:** https://opendataphilly.org/datasets/philadelphia-land-cover-raster/
- **Use:** Primary comfort indicator - shade from trees

Transit Stop Locations

- **Source:** SEPTA GTFS (General Transit Feed Specification)
- **Provider:** Southeastern Pennsylvania Transportation Authority
- **Coverage:** All SEPTA modes (bus, subway, trolley, regional rail)
- **Update Frequency:** Regular updates (last updated June 2025)
- **Access:** Free download from OpenDataPhilly.org or SEPTA Developer Portal
- **Format:** Standard GTFS format (stops.txt contains stop locations)

Secondary Data (Optional)

Street Width & Characteristics

- **Source:** OpenStreetMap (via OSMnx)
- **Access:** Extracted as street network attributes
- **Variables:** Street width, number of lanes, street type
- **Purpose:** Wider streets = more sun exposure, narrower streets = more shade from buildings

Land Cover Types

- **Source:** Philadelphia Land Cover Raster (same as tree canopy)
- **Variables:** Grass/shrub, pavement types, water features
- **Purpose:** Surface heat absorption varies by material (asphalt vs. grass)

Demographics

- **Source:** U.S. Census Bureau American Community Survey (ACS)
- **Access:** Census API via `censusdata` or `tidycensus` Python packages
- **Variables:** Income, race/ethnicity, car ownership, commute mode
- **Purpose:** Equity analysis of comfortable transit access

---

4. Methodology

Phase 1: Data Collection & Preprocessing

Street Network (Week 1)

- Extract Philadelphia walkable street network using OSMnx

- Filter to pedestrian-accessible streets
- Create network graph with edges (street segments) and nodes (intersections)

Tree Canopy Data (Week 1)

- Download Philadelphia Land Cover Raster (2018)
- Reproject to match street network coordinate system (EPSG:4326 or local projection)
- Extract tree canopy class from multi-band raster

Transit Stops (Week 1)

- Download SEPTA GTFS data
- Extract stop locations from stops.txt
- Convert to GeoDataFrame with point geometries
- Filter to frequent service stops (>10 trips/day) for analysis focus

Phase 2: Comfort Score Calculation (Week 2)

Multi-Factor Comfort Scoring

The comfort score for each street segment will be calculated using multiple environmental factors:

Primary Factor: Tree Canopy (Weight: 60%) For each street segment:

1. Create buffer polygon around street centerline (10-15 meter buffer)
2. Use `rasterstats` to calculate tree canopy coverage within buffer
3. Tree score = percentage of buffer covered by canopy (0-1)

Secondary Factor: Street Width (Weight: 20%) From OSMnx street attributes:

1. Extract street width or number of lanes
2. Width score = inverse of normalized width (narrower = higher score)
3. Rationale: Narrow streets receive more shade from adjacent buildings

Tertiary Factor: Surface Type (Weight: 20%) From land cover raster:

1. Calculate percentage of non-tree vegetation (grass, shrubs) vs. paved surfaces
2. Green score = ratio of vegetated to paved surfaces
3. Rationale: Vegetation cools through evapotranspiration, reduces heat reflection

Combined Comfort Score:

```
comfort_score = (0.6 × tree_score) + (0.2 × width_score) + (0.2 × green_score)
```

Result: Each street segment receives a comfort score from 0 (least comfortable) to 1 (most comfortable)

Quality Checks

- Visualize scores on sample street segments
- Validate against local knowledge of comfortable/uncomfortable streets
- Adjust weights if needed based on preliminary results
- Check for data gaps or anomalies

Phase 3: Route Analysis (Week 3-4)

Sample Location Selection

- Select 20-30 sample origin points distributed across Philadelphia
- Prioritize residential areas, schools, hospitals, employment centers
- Ensure geographic and demographic diversity

For Each Sample Location:

1. **Identify nearest transit stops** (3-5 closest stops)
2. **Calculate shortest distance route** to each stop
   - Standard NetworkX shortest path algorithm
   - Weight edges by distance only
   - Record: distance, estimated time, comfort score
3. **Calculate comfort-optimized route** to same stops
   - Custom routing with weighted edges: `weight = distance × (1 + comfort_penalty)`
   - Where `comfort_penalty = (1 - comfort_score) × sensitivity_factor`
   - Prioritizes streets with higher comfort scores
   - Record: distance, estimated time, comfort score
4. **Compute trade-offs**
   - Additional distance for comfort benefit

    - Percent improvement in comfort score
    - Identify routes where comfort optimization is most beneficial

Network-Wide Analysis

- Calculate comfort scores within 5, 10, 15-minute walks of all transit stops
- Identify "transit deserts" with poor service
- Identify "comfort deserts" with poor environmental quality
- Find overlap: areas with both transit and comfort challenges

Phase 4: Equity Analysis (Week 4)

Demographic Overlay (If Time Permits)

- Join Census demographic data to analysis areas
- Compare comfort access by:
    - Median household income
    - Racial/ethnic composition
    - Car ownership rates (transit dependence)
    - Age distribution (heat vulnerability, mobility limitations)
- Identify environmental justice concerns

Phase 5: Interactive Dashboard Development (Week 5)

Panel Dashboard Features

Build an interactive web application using Panel that allows users to:

Core Functionality:

1. **Address Input**
    - Geocode user-provided origin address
    - Display on map with transit stops nearby
2. **Route Calculation**
    - Find 3 nearest transit stops
    - Calculate and display:
        - Shortest distance route (baseline)

- Comfort-optimized route
- Balanced route (moderate optimization)

3. **Interactive Controls**
   - Slider: Comfort vs. Distance preference (0-100%)
   - Checkboxes: Enable/disable comfort factors (trees, width, surfaces)
   - Dropdowns: Time of day, season (if implemented)
4. **Visualizations**
   - Interactive map (Folium or Plotly) showing:
     - Color-coded street comfort scores
     - Transit stop locations
     - Multiple route options overlaid
     - Pop-ups with route details
   - Comparison charts:
     - Bar chart: Distance vs. comfort trade-offs
     - Radar chart: Comfort factor breakdown by route
     - Table: Detailed metrics for each route
5. **Deployment**
   - Deploy on Hugging Face Spaces (free, easy) or Railway.app
   - Responsive design for mobile and desktop
   - Fast load times with cached network data

Technical Implementation:

```python
```

```python
import panel as pn
import geoviews as gv
import holoviews as hv

# Pre-load comfort-scored network
# Create geocoding function
# Build routing function with adjustable weights
# Create interactive map with route overlays
# Deploy with pn.serve() or Panel server
```

Why This Adds Value:

- Makes analysis accessible to non-technical stakeholders
- Demonstrates full-stack geospatial data science skills
- Creates shareable, memorable deliverable
- Shows you can translate analysis into usable tools

Phase 6: Final Documentation & Polish (Week 6)

Interactive Maps

1. **Citywide comfort map -** Heat map of pedestrian comfort scores on streets
2. **Transit stop accessibility -** Isochrones showing walk times to stops
3. **Route comparison examples -** Side-by-side direct vs. comfort-optimized routes
4. **Priority improvement areas -** Neighborhoods needing environmental interventions

Analysis Visualizations

1. **Trade-off scatter plots -** Distance increase vs. comfort improvement
2. **Comfort factor breakdown -** Contribution of trees, street width, surface type
3. **Neighborhood comparisons -** Comfort access by area
4. **Equity charts -** Access disparities by demographics
5. **Sensitivity analysis -** How different comfort factor weights affect results

Quarto Website Structure

- **Home/Introduction -** Problem statement, pedestrian comfort factors, motivation
- **Data & Methods -** Data sources, comfort scoring methodology, routing algorithm

- **Analysis -** Route analysis results with interactive maps and charts
- **Comfort Factors -** Breakdown of what makes routes comfortable
- **Interactive Tool -** Embedded Panel dashboard or link to deployed app
- **Equity -** Demographic analysis and environmental justice findings
- **Recommendations -** Priority areas for urban improvements
- **Conclusion -** Limitations, future work, policy implications
- **Appendix -** Technical details, code snippets, additional visualizations

---

5. Expected Deliverables

Technical Deliverables

1. **GitHub Repository** containing:
    - Python notebooks for data processing and analysis
    - Scripts for shade scoring and routing
    - Processed datasets (where license permits)
    - README with project overview and website URL
2. **GitHub Repository** (hosted on GitHub Pages) with:
    - Multi-page documentation of full pipeline
    - Interactive maps created with Folium or Plotly
    - Static visualizations using matplotlib/seaborn
    - **Embedded or linked Panel dashboard**
    - Clear narrative connecting methods to findings
    - High-quality figures and professional presentation
    - Code appendix with key functions documented

Analysis Deliverables

1. **Comfort Coverage Analysis**
    - Citywide map of pedestrian comfort scores on walking network
    - Statistics on comfort levels by neighborhood

- Identification of low-comfort corridors

2. **Route Comparison Analysis**
   - 20-30 case studies showing direct vs. comfort-optimized routes
   - Quantification of distance/comfort trade-offs
   - Factor analysis: which comfort elements matter most?
   - Identification of routes where optimization is most beneficial
3. **Transit Access Analysis**
   - Walking time isochrones to transit stops
   - Coverage analysis: % of population within comfortable walk of transit
   - Gaps in service and environmental comfort
4. **Equity Assessment**
   - Comparison of comfortable transit access across demographics
   - Identification of underserved neighborhoods
   - Priority areas for intervention
5. **Policy Recommendations**
   - Specific streets/corridors for improvement priority
   - Multi-modal interventions: trees, street narrowing, surface changes
   - Estimated impact of different interventions
   - Potential for comfort-aware navigation tools
6. **Interactive Dashboard**
   - Deployed Panel application for route exploration
   - User-friendly interface for testing different scenarios
   - Real-time visualization of comfort-optimized routes
   - Demonstrates practical application of research

---

6. Project Timeline

Week 1 (Nov 6-12): Data Collection & Setup

- ✅ Proposal approval

- Download all datasets
- Test data loading and basic operations
- Set up project repository and Quarto template

Week 2 (Nov 13-19): Data Processing

- Process tree canopy raster and other land cover types
- Extract street characteristics (width, type) from network
- Develop multi-factor comfort scoring methodology
- Create comfort-scored network graph
- Test and validate scoring approach

Week 3 (Nov 20-26): Initial Analysis

- Select sample locations
- Implement routing algorithms
- Run route comparisons for subset of locations
- Create initial visualizations
- *Note: Thanksgiving break - lighter workload*

Week 4 (Nov 27-Dec 3): Complete Analysis

- Finish route analysis for all sample locations
- Conduct network-wide accessibility analysis
- Analyze contribution of different comfort factors
- Perform equity analysis (if time)
- Refine visualizations

Week 5 (Dec 4-10): Dashboard & Visualization

- Develop Panel dashboard with core functionality
- Test and debug interactive features
- Deploy dashboard to hosting platform
- Create static visualizations for Quarto site
- Begin website documentation

Week 6 (Dec 11-14): Final Polish & Integration

- Complete Quarto website with all sections
- Integrate dashboard (embed or link)
- Write comprehensive documentation
- Develop policy recommendations
- Final testing and validation
- Proofread and edit all content
- Prepare submission materials
- **Submit by Dec 14 @ 11:59 PM**

---

7. Technical Considerations

Python Libraries

Core Libraries:

- `osmnx` - Street network extraction and analysis
- `networkx` - Graph routing algorithms
- `geopandas` - Spatial data manipulation
- `rasterio` - Raster data reading/writing
- `rasterstats` - Zonal statistics for shade scoring
- `pandas` - Data manipulation
- `numpy` - Numerical operations

Visualization:

- `folium` - Interactive web maps
- `matplotlib` / `seaborn` - Static plots
- `plotly` - Interactive charts
- `contextily` - Basemaps
- `panel` - Interactive dashboard framework

- `holoviews` / `geoviews` - Panel-compatible visualizations

Optional:

- `censusdata` - Demographic data access
- `gtfs_kit` - GTFS data parsing
- `shapely` - Geometric operations

Computational Requirements

- **Processing time:** Tree canopy sampling may be computationally intensive
  - Solution: Process in batches, cache intermediate results, pre-compute comfort scores
- **Memory:** Large raster files require careful handling
  - Solution: Use windowed reading, downsample if needed, store processed scores
- **Storage:** Moderate (< 5 GB for all data and outputs)
- **Dashboard performance:** Network data must load quickly
  - Solution: Pre-process and pickle comfort-scored network, use spatial indexing

Risk Mitigation

Potential Challenge: Comfort scoring algorithm is too subjective or weights are unclear

- **Mitigation:** Review urban planning literature on pedestrian comfort factors
- **Backup plan:** Focus on tree canopy only (simpler, still valid), make weights adjustable
- **Sensitivity analysis:** Test how different weight combinations affect results

Potential Challenge: Comfort scoring takes too long computationally

- **Mitigation:** Start with small study area, optimize before scaling
- **Backup plan:** Simplify to 2-factor model (trees + street width only)

Potential Challenge: Street width data incomplete in OSM

- **Mitigation:** Use street type as proxy (residential vs. arterial)
- **Backup plan:** Focus on tree canopy and land cover (both confirmed available)

Potential Challenge: Route optimization produces unrealistic routes

- **Mitigation:** Add constraints (maximum distance multiplier, avoid highways)

- **Validation:** Compare against real pedestrian behavior patterns

---

8. Success Criteria

Minimum Viable Project (B/B+ Grade)

- ✅ Complete comfort scoring of Philadelphia street network using tree canopy
- ✅ Working route comparison tool (direct vs. comfort-optimized routes)
- ✅ Analysis of 15+ sample locations
- ✅ Clear visualizations showing trade-offs
- ✅ Professional Quarto website documenting process
- ✅ Meets 2 technical requirements (OSMnx + raster analysis)

Target Project (A-/A Grade)

- All minimum requirements PLUS:
- ✅ Multi-factor comfort scoring (trees + street width + surface type)
- ✅ Comprehensive coverage (25+ sample locations)
- ✅ Network-wide accessibility analysis
- ✅ Factor contribution analysis (what matters most for comfort?)
- ✅ Interactive visualizations
- ✅ Equity analysis with demographic data
- ✅ Actionable, evidence-based policy recommendations
- ✅ Well-written, compelling narrative

Stretch Goals (A/A+ Territory)

- Interactive Panel dashboard for route finding
- Integration of additional comfort factors (traffic noise, air quality)
- Time-of-day analysis (morning vs. afternoon sun angles)
- Validation with pedestrian preference surveys or GPS trace data
- Cost-benefit analysis of different intervention types
- Comparison across different Philadelphia neighborhoods

- Seasonal variation in comfort (summer vs. winter preferences)

---

9. Expected Outcomes & Impact

Academic Contributions

- Novel application of multi-criteria routing to pedestrian comfort
- Methodology replicable for other cities
- Demonstrates integration of raster and vector geospatial data
- Showcases practical application of weighted network optimization
- Advances understanding of pedestrian preferences beyond simple distance

Practical Applications

- **Urban planning:** Evidence-based priority-setting for street improvements
- **Public health:** Identify comfort-deprived populations and routes
- **Transportation:** Inform pedestrian infrastructure and transit access investments
- **Technology:** Foundation for comfort-aware navigation apps (like "CoolWalks")
- **Equity:** Highlight and address environmental justice issues
- **Climate adaptation:** Multi-benefit street interventions (comfort + resilience)

Portfolio Value

- Demonstrates full data science pipeline
- Shows proficiency in geospatial analysis
- Addresses timely, relevant urban challenge
- Strong visual component for presentations
- Applicable to urban planning, transportation, climate, and tech careers

---

10. Conclusion

This project addresses pedestrian comfort—a critical yet underexplored dimension of urban mobility—at the intersection of climate change, public health, and transportation equity. By moving beyond simple shade analysis to a multi-factor comfort framework, it provides a more nuanced understanding of what makes streets pleasant and accessible for walking.

The project scope is ambitious yet achievable for solo work, with clear minimum requirements and defined stretch goals. All necessary data has been confirmed available and accessible. The methodology is sound, building on established techniques in network analysis and raster processing while introducing novel comfort-based routing algorithms.

Most importantly, this project produces actionable insights that can directly improve pedestrian experiences in Philadelphia. The multi-factor approach allows for diverse interventions—tree planting, street narrowing, surface improvements—giving planners flexibility to match solutions to local contexts and budgets. By identifying where comfort-optimized routes diverge significantly from direct routes, the analysis highlights priority corridors where environmental improvements would have the greatest impact on daily pedestrian experiences.

I am excited to pursue this project and confident in its feasibility, rigor, and real-world applicability.

References & Data Links

Data Sources:

- Philadelphia Land Cover Raster: https://opendataphilly.org/datasets/philadelphia-land-cover-raster/
- SEPTA GTFS: https://opendataphilly.org/datasets/septa-gtfs/
- OpenStreetMap (via OSMnx): https://osmnx.readthedocs.io/

Technical Documentation:

- OSMnx Documentation: https://osmnx.readthedocs.io/
- Rasterio Documentation: https://rasterio.readthedocs.io/
- Rasterstats Documentation: https://pythonhosted.org/rasterstats/

Background Research:

- Urban heat island effect and tree canopy studies
- Last mile challenges in public transportation
- Environmental justice in urban planning