Retail Trade-Area & Network Planning Using ArcGIS, GeoPandas, and Power BI

Abstract

This project develops an end-to-end geospatial decision-support pipeline for retail network optimization. Using *ArcGIS Pro*, *GeoPandas*, and *Power BI*, it integrates store and competitor locations with Statistics Canada 2021 Census data to analyze population reach, drive-time coverage, and market potential. The workflow builds drive-time isochrones, enriches each zone with demographic indicators, and applies a **Huff spatial interaction model** to estimate demand probabilities and branch-level market shares. The outputs feed directly into a Power BI dashboard, enabling managers to visualize demographic catchments, branch dominance, and cannibalization across Ontario. The system demonstrates how open geospatial and demographic data can guide evidence-based site selection and consolidation decisions.

1. Introduction

Retail chains must regularly assess how well their branch network matches local demand. Traditional site-selection approaches rely on static buffers or sales data, which overlook demographic structure and competitor proximity.

This project introduces a reproducible, data-driven method for trade-area delineation and demand modeling. The objective was to design a spatially intelligent dashboard capable of answering three key questions:

1. **Coverage:** Which population segments fall within each branch's 5-, 10-, and 15-minute drive times?
2. **Attractiveness:** How do income and household density influence branch performance?
3. **Competition:** Which branches dominate specific demand zones under a Huff-model probability framework?

2. Data Sources

Dataset	Description	Source
2021 Census Profile (DA-level)	Population, households, and median income for dissemination areas	Statistics Canada (98-401-X2021006)
DA Boundary Shapefile	Geographic boundaries for dissemination areas (Ontario)	StatCan 2021 Digital Boundary Files
Branch & Competitor POIs	Latitude/longitude of sample retail stores	Manually generated in ArcGIS Pro
Ontario Road Network – Composite Service	Routing and drive-time computation	ArcGIS Online / Esri Living Atlas
Derived Datasets	Isochrones, demographic joins, Huff results	Created via GeoPandas + ArcGIS Pro

Table 1: Dataset Sources & Description

3. Methodology

3.1 Data Integration and Pre-Processing

- Census tables were chunk-read in *pandas* (≈ 8 GB CSV) and filtered to "Dissemination Area" rows.
- Key fields: population, households, and median income, were cleaned and joined to DA geometries in *GeoPandas* and *ArcGIS Pro*.
- Branch points were projected to WGS 84; centroids of each DA served as demand origins.

3.2 Drive-Time Isochrones

ArcGIS Pro's *Network Analyst* → *Service Area* tool used the Ontario Road Network to generate 5-, 10-, 15-minute polygons for each branch.

Demographic data were spatially joined to these polygons to calculate population, households, and income totals per band. Finally, forming the **FactTradeArea** table later consumed in Power BI.

3.3 Demand Reduction and Clustering

Because Ontario includes >100 000 DAs, the project applied *K-Means Cluster Analysis* (K ≈ 500) on DA centroids using population and income as features.

For each cluster, a **population-weighted mean center** was computed, producing ≤ 1 000 representative **demand centers** with total cluster population attributes.

3.4 Huff Model Implementation

Using Python (GeoPandas + NumPy), I implemented a gravity-based Huff model:

$$P_{ij} = \frac{A_j / D_{ij}^{\beta}}{\sum_k A_k / D_{ik}^{\beta}}$$

where
P_{ij} = probability that demand i chooses branch j,

A_j = branch attractiveness (e.g., median income or composite index),

D_{ij} = drive distance (km),

β = distance-decay parameter (≈ 1.8).

The script computed:

- Pairwise haversine distances (demand → branch)
- Normalized probabilities
- Expected customers (E_{ij}=P_{ij} \times \text{population}_i)
- Branch-level market share and dominant-branch zones

Outputs:

- huff_od_probabilities.csv
- huff_market_share.csv
- huff_demand_dominant.geojson

4. Visualization and Decision Dashboard

4.1 Power BI Model

Tables integrated into the data model:

Table	Key Fields
DimFacility	BranchID, Branch Name, City, Coords
DimBand	BandLabel (5–10–15 min)
FactTradeArea	FacilityID, ToBreak, Demographics
FactHuff_OD	OriginID, BranchID, prob, expected_cust
FactHuff_Branch	BranchID, expected_cust, market_share_%
GeoHuff_Demand	CLUSTER_ID, DomBranch, DomProb, PopWeighted

Table 2: Data Model Tables

Relationships: FactHuff_Branch[BranchID] → DimFacility[BranchID];

FactHuff_OD[OriginID] → GeoHuff_Demand[CLUSTER_ID].

4.2 Dashboard Pages

(Page 1) Network Overview

(1.1) Analyze using Census

– KPI cards (population, households, median income, area, density) plus Huff metrics (expected customers, market share).

(1.2) Analyze using the Huff Model– Comparative charts:

- *Census vs Huff Expected Population by Branch*
- *Attractiveness vs Market Share scatter plot*
- Cannibalization Index gauge (1 – mean max probability).

(Page 2) Map

(2.1) Analyze using Census

ArcGIS Maps for Power BI displays different time-driving zone trade-area polygons, branch bubbles sized by population and median income.

(2.2) Analyze using the Huff Model

ArcGIS Maps for Power BI displays different time-driving zone trade-area polygons, branch points colored by modeled share, and sized by modeled expected customers.

All visuals share synchronized slicers for City, Band, Branch, and optional β scenarios.

5. Results and Insights

- **Model Alignment:** Huff-predicted customer totals correlated strongly ($r \approx 0.91$) with demographic population counts, validating distance decay selection.
- **Dominant Catchments:** Urban branches captured >70 % of modeled probability within 10 km, while outer suburban stores exhibited overlapping zones and potential redundancy.
- **Attractiveness Sensitivity:** When attractiveness (A_j) was weighted by median income, higher-income clusters shifted demand toward premium branches, reducing overlap.
- **Decision Support:** The dashboard enables regional managers to identify underserved clusters (low probability, high population) and simulate network consolidation scenarios without additional field surveys.

6. Conclusion

This project demonstrates a complete spatial-analytics pipeline integrating open data, ArcGIS services, and Power BI.

By combining **geospatial modeling** and **probabilistic demand estimation**, retailers can move beyond static trade-area maps toward dynamic, data-driven market intelligence.
Future work could incorporate competitor attractiveness, temporal traffic data, and automated β-calibration to further refine location-allocation recommendations.

7. Tools & Technologies

Category	**Tools**
Spatial Processing	ArcGIS Pro (Network Analyst), GeoPandas, Shapely
Data Handling	Python (pandas, NumPy)
Visualization & BI	Power BI (ArcGIS Maps visual, Shape Map)
Data Sources	Statistics Canada 2021 Census, Ontario Road Network, ArcGIS Online