Unraveling Seattle's Travel Patterns: A Data Science Exploration

**Introduction:**

- Brief overview of the significance of understanding travel patterns in urban areas like Seattle.

- Introduction to the project's objectives: to analyze social demographic factors, origin-destination pairs, transportation modes, and predict travel behaviors.

- Highlight the interdisciplinary nature of the project, combining data science, transportation planning, and urban geography.

**Methods:**

1. Data Collection and Preprocessing:

- Explanation of data sources including PSRC for demographic factors, GTFS for transit data, and Seattle Geography data for spatial features.

- Preprocessing steps such as cleaning, feature engineering, and geospatial analysis.

2. Model Training and Analysis:

- Description of ML models used (Random Forest, Linear Regression, Gradient Boosting, Deep Learning) and their suitability for predicting travel volume based on demographic factors.

- Explanation of network analysis techniques (Centralities, Pagerank, PerronRank, HodgeRank) for analyzing origin-destination pairs and predicting travel trails.

- Explore algebraic topology, topological data analysis on the network flow analysis. Innovate new clustering method on directed flow graph based on Hodge-Kondaira decomposition

- Overview of Bayesian methods for predicting transportation mode choice based on conditional probabilities.

**Results:**

1. Demographic Analysis:

- Presentation of findings on how social demographic factors influence travel behavior.

- Visualization of model predictions and their accuracy in predicting travel volume for different demographic groups.

2. Network Analysis:

- Visualization of the travel network between census tracts.

- Identification of influential nodes and routes using PageRank, PerronRank and HodgeRank.

- New clustering methods on directed flow network based on Hodge-Kondaira decomposition.

3. Transportation Mode Prediction:

- Bayesian model results showing probabilities of using different transportation modes for trips between census tracts.

- Comparison of predicted mode choices against ground truth data.

**Discussion:**

- Interpretation of findings and their implications for transportation planning in Seattle.

- Discussion of limitations including data biases, model assumptions, and uncertainties.

- Suggestions for future research including incorporating real-time data and dynamic modeling approaches.

- Comparison of different clustering and ranking methods, especially in the setting of directed flow graph

- Conclusion highlighting the project's contributions to understanding travel behavior and its potential for informing policy and infrastructure decisions in Seattle.

