SmartHop

Data Assessment Document

Members: Tanveer | Anshika | Vishal | Yousif | Saswati | Vaishnav |

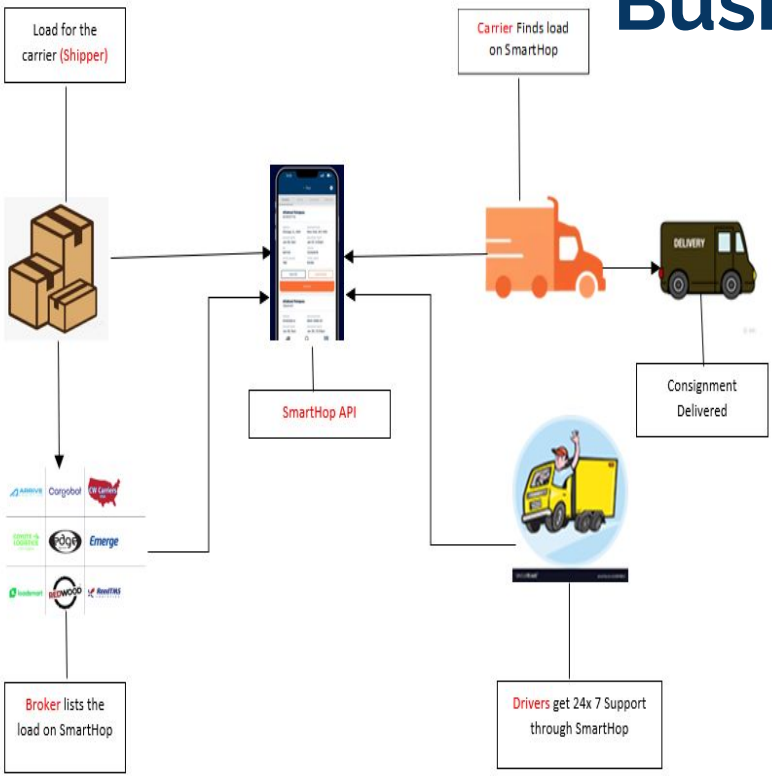
University of Maryland Robert H. Smith School of Business

Agenda

Objectives

- Business Problem and Objective
- Data Source
- Data Cleaning
- Data Preparation
- Model Considerations
- Feature Engineering
- Preliminary Findings
- Project Success Assessment

Business Problem

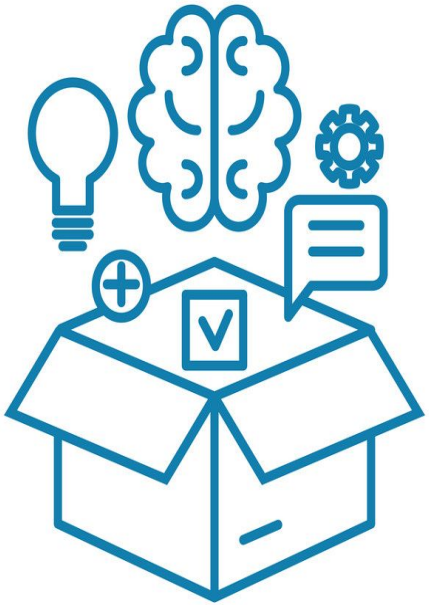


- Unclear impact of carrier's geographical location and risk appetite impacts SmartHop's performance:
  - Market Volatility in terms of volume
  - Varying profitability based on preference in load selection and markets

Objective

- Recommend strategies - routes, load choice or carrier behavior to improve the performance of carriers based on geographical analysis.

Data Source



We obtained the data from Smarthop's database on Snowflake.

The tables used are as following:
BOOKED_TRIPS
CARRIERS
CLUSTER_INFO
DRIVERS
TRUCKS
LOADS
Estimated Cost Structure Excel sheet

SmartHop

Data Cleaning

Data pre-processing

- Size of our primary table (i.e Booked_Trips) is above 7.3 MB with more than 47,000 entries.
- Load data - threshold and extraction.
- We used data of four years for our analysis (2019, 2020, 2021, 2022).
- Deduplicated the load data reducing the volume from 57M+ to 17.5M.
- Dropping cancelled trips from our analysis.
- Major analysis is done for origin and destination clusters and dropped all the NA in those columns.
- Considering only those trucks and drivers that have a date of entry recorded into the ecosystem.

Data Preparation

- New column added - Trip_Cost calculated for each booked trip from Estimated trip cost dataset to subsequently calculate Profit (our main independent variable).

- Populated location state from zip codes for extensive geographical analysis.

- Merged the truckers and drivers dataset on the CARRIER_OBJ_ID and DRIVER_OBJ_ID and used that concatenated dataset which is more suitable towards meeting the objectives.

Model Considerations

🚚 **Network & cluster analysis**

- Identify carriers' routes with profitability and cluster volume

💰 **Dead market premium analysis**

- Analyze dead market clusters for profit opportunity

↻ **Behavior analysis**

- Correlate profit to relevant carriers' behavior and incorporate the insights into recommendations

SmartHop

Feature Engineering

Identify carrier routes and profitability → Analyze dead markets → Identify correlations to carriers behavior → Incorporate insight into recommendation

- **Network analysis**
Routes and profitability of the carriers' booked trips

- **Cluster analysis**
Profitability of carriers based on cluster volume.

- **Premium analysis**
Analysis of inherit premium or cost associated to dead market

- **Behavior analysis**
Compute the correlations between identified features to interpret carriers' behaviors

- **Profit analysis and recommendations**
Summarize the accumulated insights to improve the profitability of carriers

SmartHop

Preliminary Findings

TOP 5 routes for OCLUSTER with MAX trips by main_city_name:

1. Charlotte-Atlanta
2. Columbia-Pittsburgh
3. Dallas-Philadelphia
4. Austin-Austin
5. Philadelphia-Miami

TOP 5 routes for DCLUSTER with MAX trips by main_city_name:

1. Atlanta-Philadelphia
2. Chicago-Minneapolis
3. Philadelphia-Miami
4. Dallas-New Orleans
5. Austin-Austin

Project Success Assessment

- Business Problem and Objective is well defined.
- Our dataset is cleaned and prepared for further analysis.
- In order to get to profitability, we have used the Cost table and created a cost column in the Booked_Trips table.
- Using this cost column, subsequently calculated the profitability for each row.
- Further, we have found Top 5 routes for both the origin and destination clusters on the basis of maximum trips made.
- Also, incorporating external data - historical fuel prices to better calculate costs for accurate profits.

Thanks