Freight Management: Predict the cost of shipment Data Understanding

Data Needed:

The following data is needed for the model:

- Order details (Order number, type, date, etc.)
- Item(s) to be shipped details (number, type, quantity, descriptions, etc.)
- Source, Destination
- Other costs (insurance, special packaging, loading/unloading, fuel costs, etc.)
- Actual Freight cost

Data Given:

The following data has been given for building the model to predict the cost of shipment:

File Name	Details of the Content/Data
equipmentcodes.csv	Contains the shipment orders details
marketzips.csv	Contains the mapping of equipment type code to description
regionlookups.csv	Contains the mapping of zip code to market
spotfreightdata.csv	Contains the mapping of zip code to region

"spotfreightdata.csv" file content details:

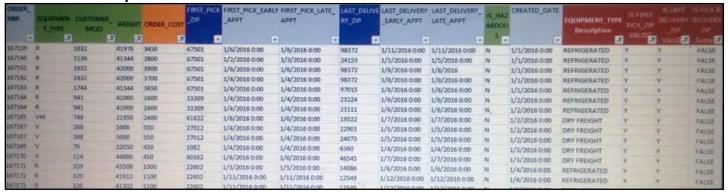
Field Name	Field Details
ORDER_NBR	Order Number
EQUIPMENT_TYPE	Equipment type code (or abbreviation). This code helps in identifying the type of
	the equipment in the shipment and useful in assessing the needful for any special
	packaging, care during transportation, etc. and would have an impact on the
	overall shipment cost
CUSTOMER_MILES	Distance between source and destination (in miles)
WEIGHT	Weight of the shipment
ORDER_COST	Shipment order cost (i.e. the target variable)
FIRST_PICK_ZIP	Pick up (or source) zip code
FIRST_PICK_EARLY_APPT	Pick up time slot (earliest)
FIRST_PICK_LATE_APPT	Pick up time slot (max by)
LAST_DELIVERY_ZIP	Delivery (or destination) zip code
LAST_DELIVERY_EARLY_APPT	Delivery time slot (earliest)
LAST_DELIVERY_LATE_APPT	Delivery time slot (max by)
IS_HAZARDOUS	does shipment or package(s) contains any hazardous item(s) or not. This is an
	important attribute and useful in assessing the need for any special handling and in
	turn would have an impact on overall shipment cost
CREATED_DATE	Ordered Date

Observation / Notes:

•	No. of records (total count)	: 37699
•	No.of records with Unique Order number	: 25631
•	No. of records with the same order number (duplicate records)	: 12068
•	No. of records with '0' or 'NA' as "CUSTOMER_MILES"	: 655
•	No. of records with blank "EQUIPMENT_TYPE"	: 2101
•	No. of records with '0' or 'NA' as "WEIGHT"	: 3506
•	No. of records with '0' or 'NA' as "ORDER_COST"	7734
•	No. of records with invalid "FIRST_PICK_ZIP"	: 200
•	No. of records with invalid "LAST_DELIVERY_ZIP"	: 326
•	No. of records for which equipment meaning not found	: 89
•	No. of records (post removal of above invalid records) left	: 25468

- To find the mapped market code, use the first 2 or 3 digits of the pickup or delivery zip code and check for the corresponding market code in "marketzips.csv"
- To find the mapped region, use the first 2 or 3 digits of the pickup or delivery zip code and check for the corresponding region in "regionlookups.csv"

Filtered data sample view:



"equipmentcodes.csv" file content details:

Field Name	Field Details
EQUIPMENT_TYPE	Equipment type code (for second level category)
new abbr	Equipment type code (for first level category)
meaning	Description of the equipment type

Observation:

No. of records with unique mapping of EQUIPMENT_TYPE to meaning : 293 No. of mappings not found in this mapping table : 7 (IMDL, SB, DRAY, VCAR, D-LTL, VS, FT-4)

"marketzips.csv" file content details:

Field Name	Field Details
Zips	Zip codes (the first 2 or 3 digits of the zip code)
Market	The mapping of the zip to the market

Observation:

No. of records with unique mapping of EQUIPMENT TYPE to meaning : 905 No. of mappings not found in this mapping table : 135

"regionlookups.csv" file content details:

Field Name	Field Details
Zip Code	Zip code (the first 2 or 3 digits of the zip code)
Region	Contains the region corresponding to the zip code

Observation:

No. of records with unique mapping of Region to zip code : 906

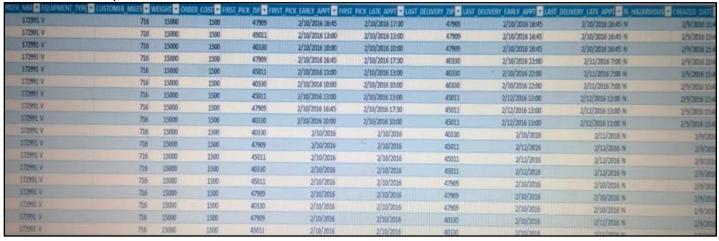
No. of unique regions found : 15

Open questions:

spotfreightdata.csv:

- ➤ Contains multiple records with the same order number, equipment type, miles, weight, order cost details but with different first pick up info (i.e. zip, early appt, late appt), last delivery info (i.e. zip, early appt, late appt) and order create date.
 - Are these records about source and destination related updates to the original order record or something else?
 - If these are just the updates to the original order record; then, which one to be considered as the latest one and/or how to identify it (means, any particular field or value to check for the same)

For example:



marketzips.csv:

- What are these zips and market codes mapping about?
- What is its relation to the shipping cost?

regionlookups.csv:

- What are these zips codes (2 digit and 3 digit) about?
- What is the relation of the region to shipping cost calculation?

References:

https://www.melissadata.com/lookups/index.htm

http://distancecheck.com

https://www.melissadata.com/lookups/zipdistance.asp