

# 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 ( <b>i.e. the target variable</b> )
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:

ORDER_NBR	EQUIPMENT_TYPE	CUSTOMER_MILES	WEIGHT	ORDER_COST	FIRST_PICK_ZIP	FIRST_PICK_EARLY_APPT	FIRST_PICK_LATE_APPT	LAST_DELIVERY_ZIP	LAST_DELIVERY_EARLY_APPT	LAST_DELIVERY_LATE_APPT	IS_HAZARDOUS	CREATED_DATE	EQUIPMENT_TYPE_Description	IS FIRST PICK_VALID	IS LAST DELIVERY_VALID	IS PICK & DELIVERY Same
167159	R	1832	41976	3450	67501	1/6/2016 0:00	1/6/2016 0:00	98372	1/11/2016 0:00	1/11/2016 0:00	N	1/1/2016 0:00	REFRIGERATED	Y	Y	FALSE
167160	R	1136	41344	2800	67501	1/2/2016 0:00	1/3/2016 0:00	24153	1/3/2016 0:00	1/5/2016 0:00	N	1/1/2016 0:00	REFRIGERATED	Y	Y	FALSE
167161	R	1832	42000	3900	67501	1/6/2016 0:00	1/6/2016 0:00	98372	1/8/2016 0:00	1/8/2016 0:00	N	1/1/2016 0:00	REFRIGERATED	Y	Y	FALSE
167162	R	1832	42000	3700	67501	1/4/2016 0:00	1/4/2016 0:00	98372	1/8/2016 0:00	1/8/2016 0:00	N	1/1/2016 0:00	REFRIGERATED	Y	Y	FALSE
167163	R	1744	41344	3650	67501	1/4/2016 0:00	1/4/2016 0:00	97015	1/8/2016 0:00	1/8/2016 0:00	N	1/1/2016 0:00	REFRIGERATED	Y	Y	FALSE
167164	R	941	41000	1600	33309	1/4/2016 0:00	1/4/2016 0:00	23224	1/6/2016 0:00	1/6/2016 0:00	N	1/2/2016 0:00	REFRIGERATED	Y	Y	FALSE
167164	R	941	41000	1600	33309	1/4/2016 0:00	1/4/2016 0:00	23111	1/6/2016 0:00	1/6/2016 0:00	N	1/2/2016 0:00	REFRIGERATED	Y	Y	FALSE
167165	VM	748	21950	2400	61822	1/6/2016 0:00	1/6/2016 0:00	19522	1/7/2016 0:00	1/7/2016 0:00	N	1/2/2016 0:00	DRY FREIGHT	Y	Y	FALSE
167167	V	268	5000	550	27012	1/4/2016 0:00	1/4/2016 0:00	22901	1/5/2016 0:00	1/5/2016 0:00	N	1/2/2016 0:00	DRY FREIGHT	Y	Y	FALSE
167167	V	268	5000	550	27012	1/4/2016 0:00	1/4/2016 0:00	24073	1/5/2016 0:00	1/5/2016 0:00	N	1/2/2016 0:00	DRY FREIGHT	Y	Y	FALSE
167169	V	79	22050	450	1082	1/4/2016 0:00	1/4/2016 0:00	6360	1/4/2016 0:00	1/4/2016 0:00	N	1/3/2016 0:00	DRY FREIGHT	Y	Y	FALSE
167170	V	114	44000	450	80162	1/6/2016 0:00	1/6/2016 0:00	46545	1/7/2016 0:00	1/7/2016 0:00	N	1/3/2016 0:00	DRY FREIGHT	Y	Y	FALSE
167171	R	329	43508	1000	22602	1/5/2016 0:00	1/5/2016 0:00	14086	1/6/2016 0:00	1/6/2016 0:00	N	1/4/2016 0:00	REFRIGERATED	Y	Y	FALSE
167172	R	320	41912	1100	22602	1/11/2016 0:00	1/11/2016 0:00	12549	1/12/2016 0:00	1/12/2016 0:00	N	1/4/2016 0:00	REFRIGERATED	Y	Y	FALSE
167173	R	320	41302	1100	22602	1/11/2016 0:00	1/11/2016 0:00	13540	1/12/2016 0:00	1/12/2016 0:00	N	1/4/2016 0:00	REFRIGERATED	Y	Y	FALSE

#### “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:

ORDER_NUM	EQUIPMENT_TYPE	CUSTOMER_MILES	WEIGHT	ORDER_COST	FIRST_PICK_ZIP	FIRST_PICK_EARLY_APPT	FIRST_PICK_LATE_APPT	LAST_DELIVERY_ZIP	LAST_DELIVERY_EARLY_APPT	LAST_DELIVERY_LATE_APPT	IS_HAZARDOUS	CREATED_DATE
172991 V		716	15000	1500	47909	2/10/2016 16:45	2/10/2016 17:30	47909	2/10/2016 16:45	2/10/2016 16:45 N		2/9/2016 15:4
172991 V		716	15000	1500	45011	2/10/2016 13:00	2/10/2016 13:00	47909	2/10/2016 16:45	2/10/2016 16:45 N		2/9/2016 15:4
172991 V		716	15000	1500	40330	2/10/2016 10:00	2/10/2016 10:00	47909	2/10/2016 16:45	2/10/2016 16:45 N		2/9/2016 15:4
172991 V		716	15000	1500	47909	2/10/2016 16:45	2/10/2016 17:30	40330	2/10/2016 22:00	2/11/2016 7:00 N		2/9/2016 15:4
172991 V		716	15000	1500	45011	2/10/2016 13:00	2/10/2016 13:00	40330	2/10/2016 22:00	2/11/2016 7:00 N		2/9/2016 15:4
172991 V		716	15000	1500	40330	2/10/2016 10:00	2/10/2016 10:00	40330	2/10/2016 22:00	2/11/2016 7:00 N		2/9/2016 15:4
172991 V		716	15000	1500	45011	2/10/2016 13:00	2/10/2016 13:00	45011	2/12/2016 13:00	2/12/2016 13:00 N		2/9/2016 15:4
172991 V		716	15000	1500	47909	2/10/2016 16:45	2/10/2016 17:30	45011	2/12/2016 13:00	2/12/2016 13:00 N		2/9/2016 15:4
172991 V		716	15000	1500	40330	2/10/2016 10:00	2/10/2016 10:00	45011	2/12/2016 13:00	2/12/2016 13:00 N		2/9/2016 15:4
172991 V		716	15000	1500	40330	2/10/2016	2/10/2016	40330	2/10/2016	2/11/2016 N		2/9/2016
172991 V		716	15000	1500	47909	2/10/2016	2/10/2016	45011	2/10/2016	2/12/2016 N		2/9/2016
172991 V		716	15000	1500	45011	2/10/2016	2/10/2016	45011	2/12/2016	2/12/2016 N		2/9/2016
172991 V		716	15000	1500	40330	2/10/2016	2/10/2016	45011	2/12/2016	2/12/2016 N		2/9/2016
172991 V		716	15000	1500	45011	2/10/2016	2/10/2016	47909	2/10/2016	2/10/2016 N		2/9/2016
172991 V		716	15000	1500	47909	2/10/2016	2/10/2016	47909	2/10/2016	2/10/2016 N		2/9/2016
172991 V		716	15000	1500	40330	2/10/2016	2/10/2016	47909	2/10/2016	2/10/2016 N		2/9/2016
172991 V		716	15000	1500	47909	2/10/2016	2/10/2016	40330	2/10/2016	2/11/2016 N		2/9/2016
172991 V		716	15000	1500	45011	2/10/2016	2/10/2016	40330	2/10/2016	2/11/2016 N		2/9/2016

### 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>