

# Solar Turbines

## *Data Quality Report*

MSBA Capstone Project  
Deliverable 1

*Prepared by*  
**Sam Angeles, Daniel Atlan, Betty Hagos,  
Ben Keefer, Tony Lantiere**

# Agenda

## Quality Assessment

- Parcel
- Accessorial
- EDI
- Freight

## Evaluations:

- ☐ Missing Values / Omissions
- ☐ Extreme Values
- ☐ Plausibility
- ☐ Consistency

## Data Cleaning Changelog

## Preliminary EDA & Next Steps

Parcel

# Parcel Overview

## Dataset statistics

Number of variables	23
Number of observations	490962
Missing cells	980843
Missing cells (%)	8.7%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	86.2 MiB
Average record size in memory	184.0 B

## Variable types

Numeric	9
Categorical	3
DateTime	3
Text	6
Unsupported	2

# Parcel Variables

Unnamed:	0
ACCESSORIAL_CHARGE	0
CARRIER_NAME2	0
DELIVERY_DATE	24481
DESTINATION_CITY	9362
DESTINATION_COUNTRY	1714
DESTINATION_RECEIVER_NAME2	21624
DESTINATION_ZIP	9393
DUTIES_AND_TAXES_CHARGES	0
HEIGHT	294158
LENGTH	294158
MODE	0

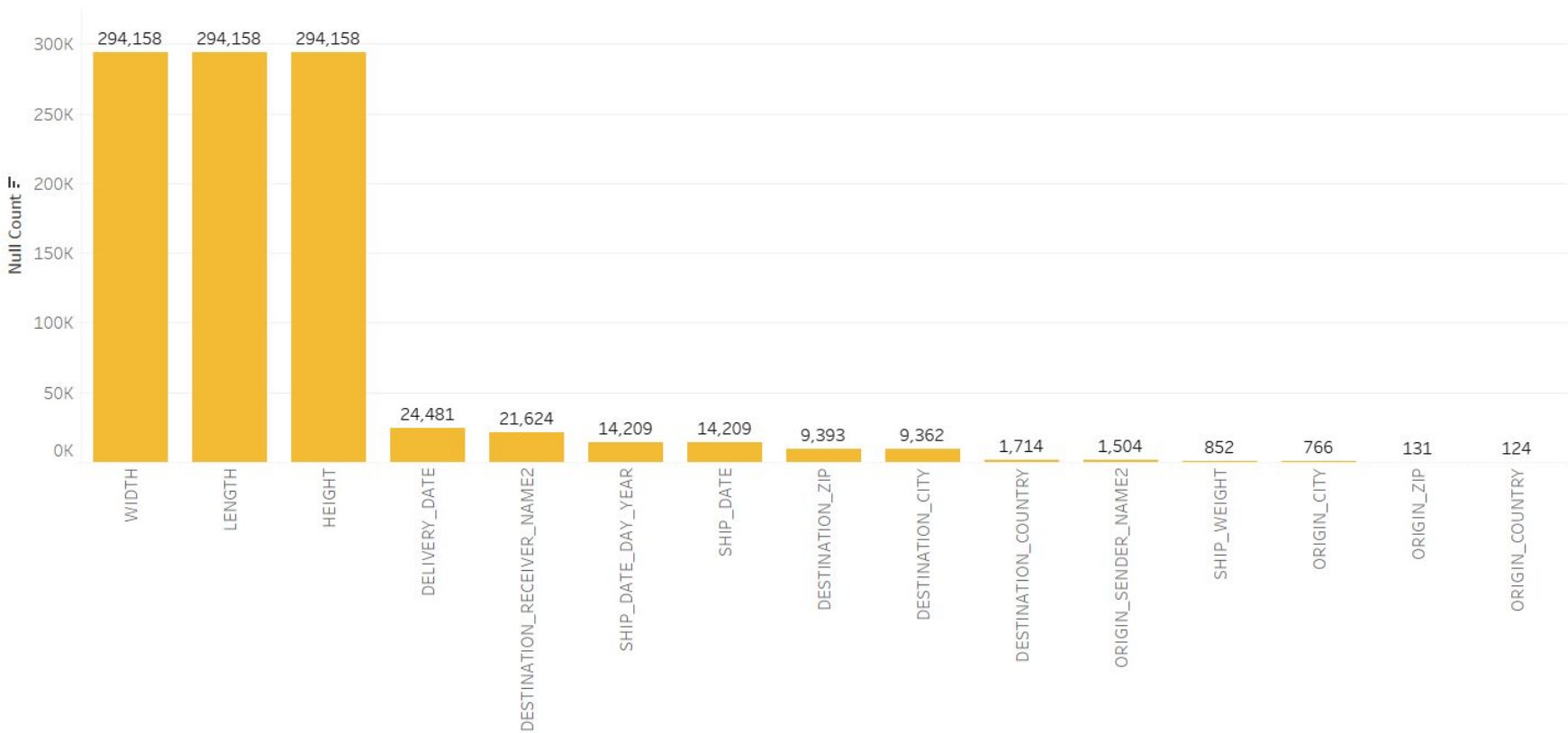
ORIGIN_CITY	766
ORIGIN_COUNTRY	124
ORIGIN_SENDER_NAME2	1504
ORIGIN_ZIP	131
SHIP_DATE	14209
SHIP_DATE_DAY_YEAR	14209
SHIP_WEIGHT	852
SHIP_WEIGHT_UNIT_OF_MEASURE	0
TOTAL_PAID	0
TRACKING_NUMBER2	0
WIDTH	294158

Count: 490,962

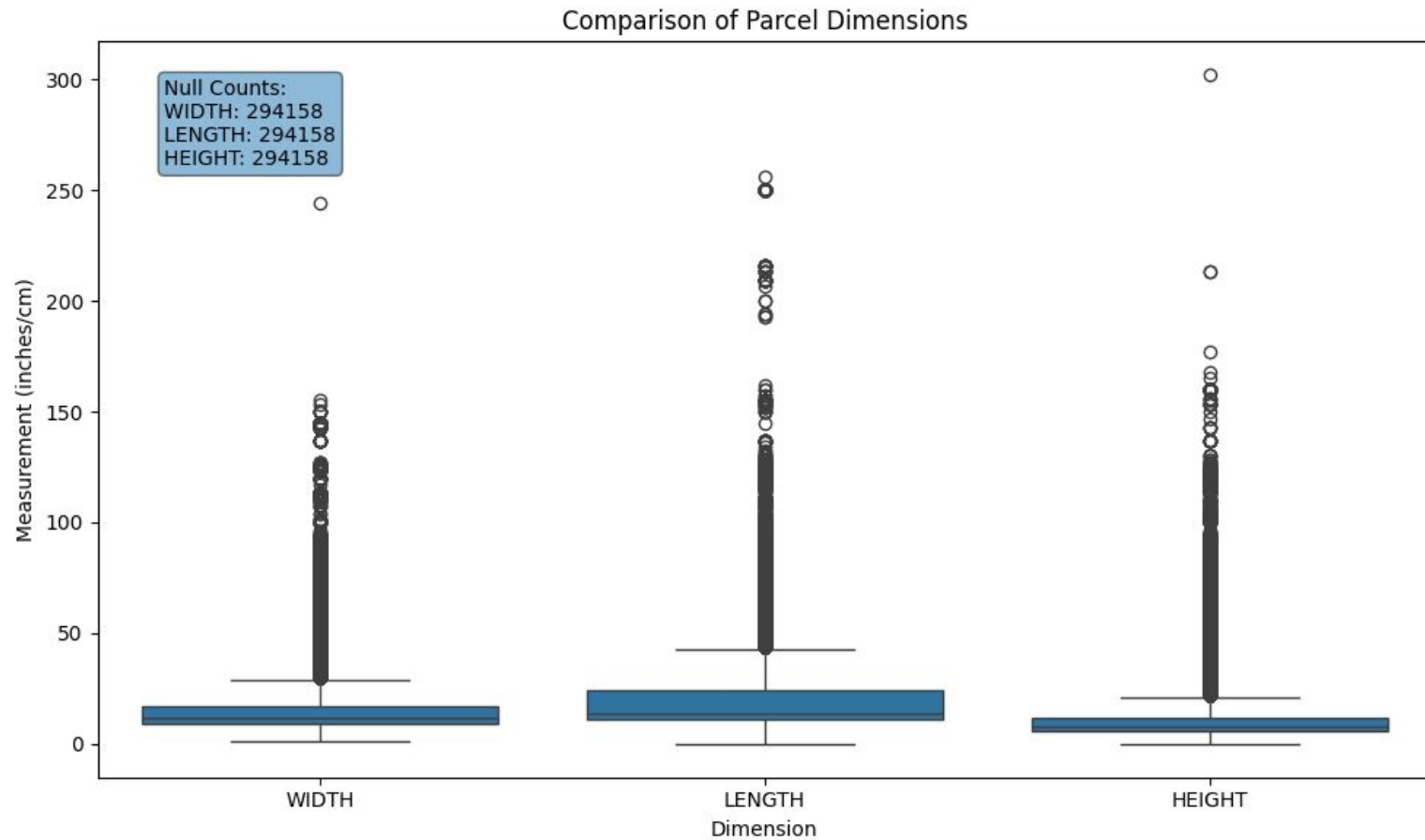
# Parcel

Null Count in *Parcel*

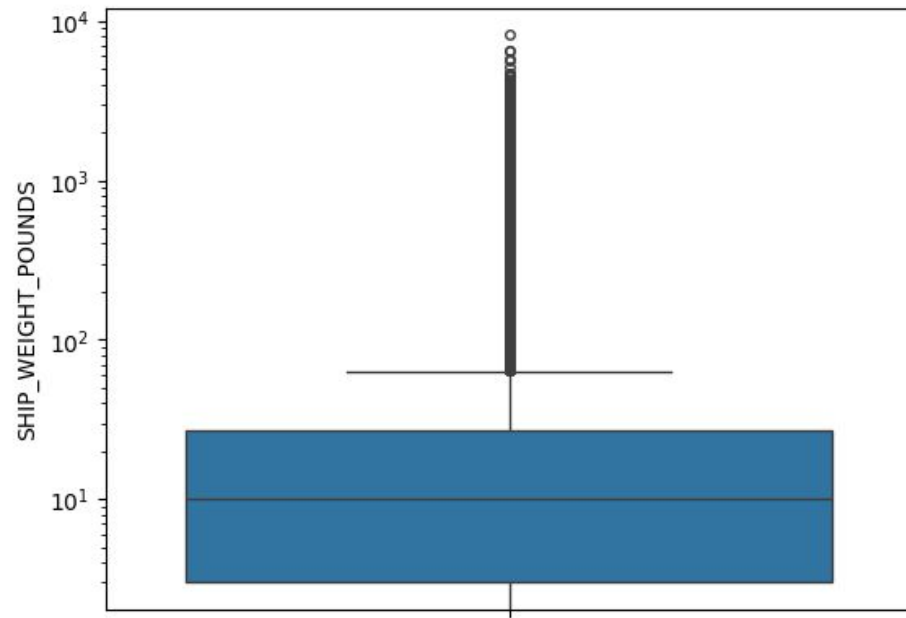
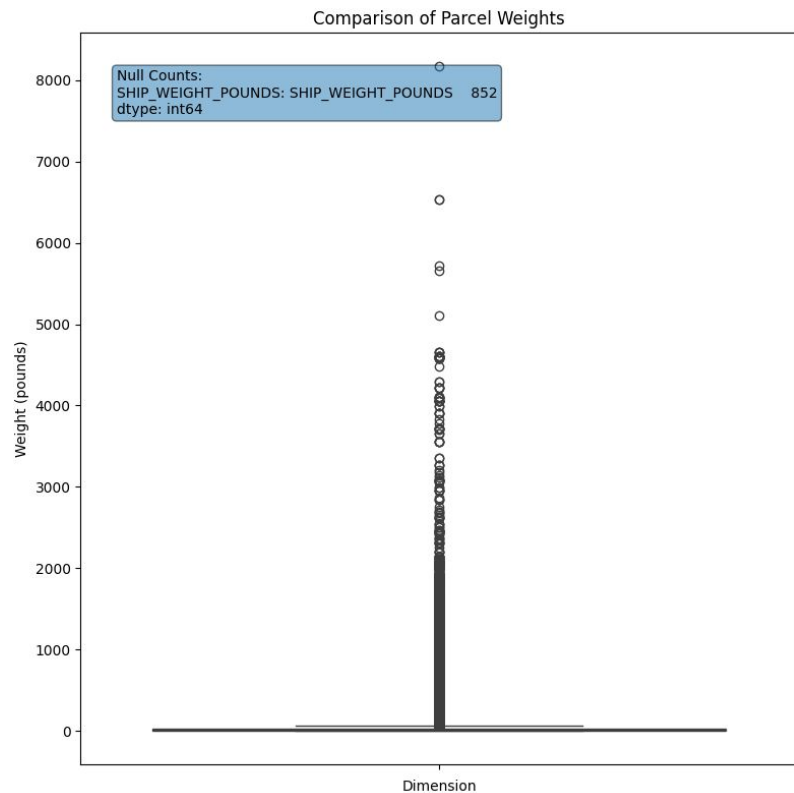
n= 490,962



# Parcel



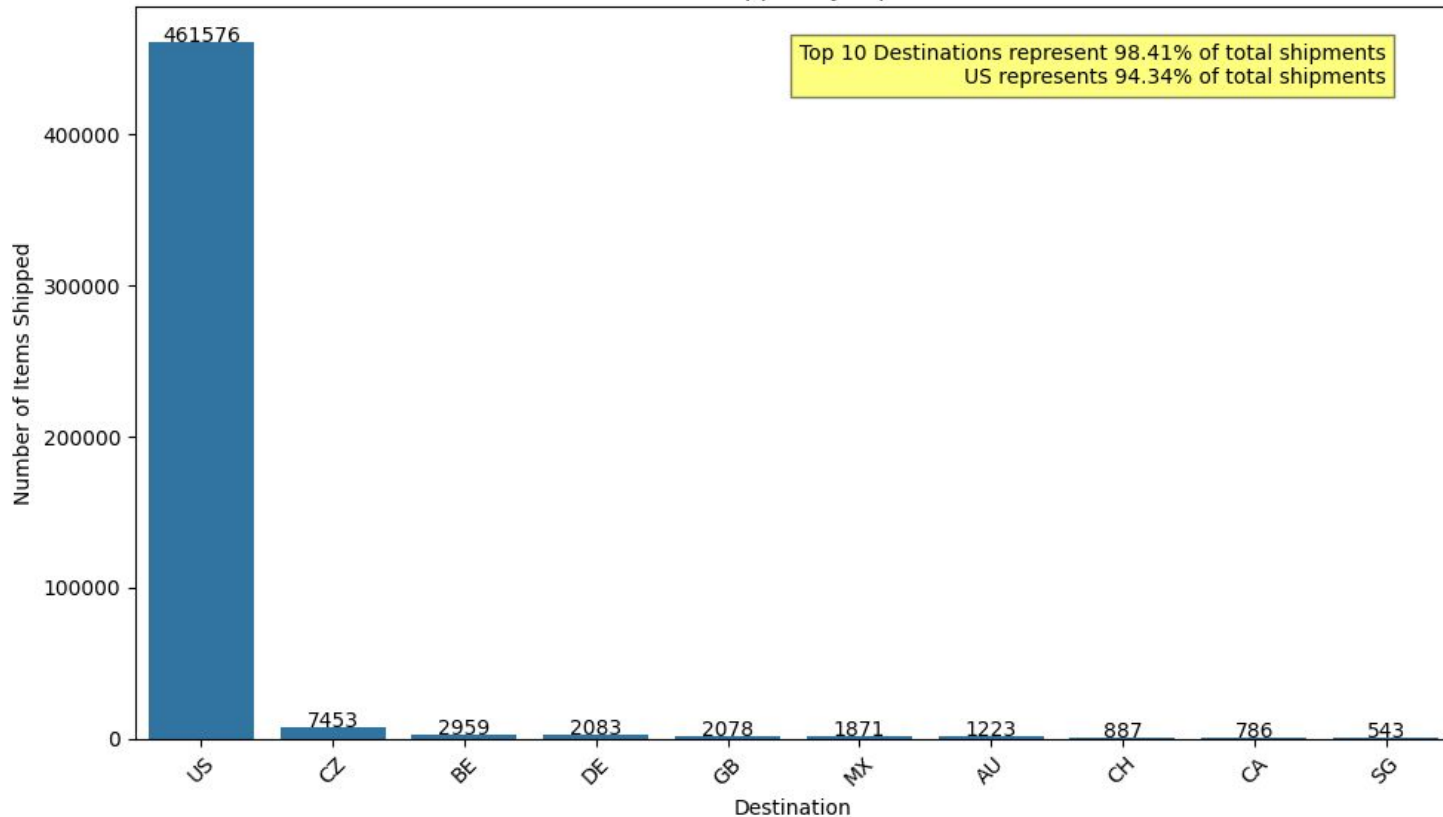
# Parcel



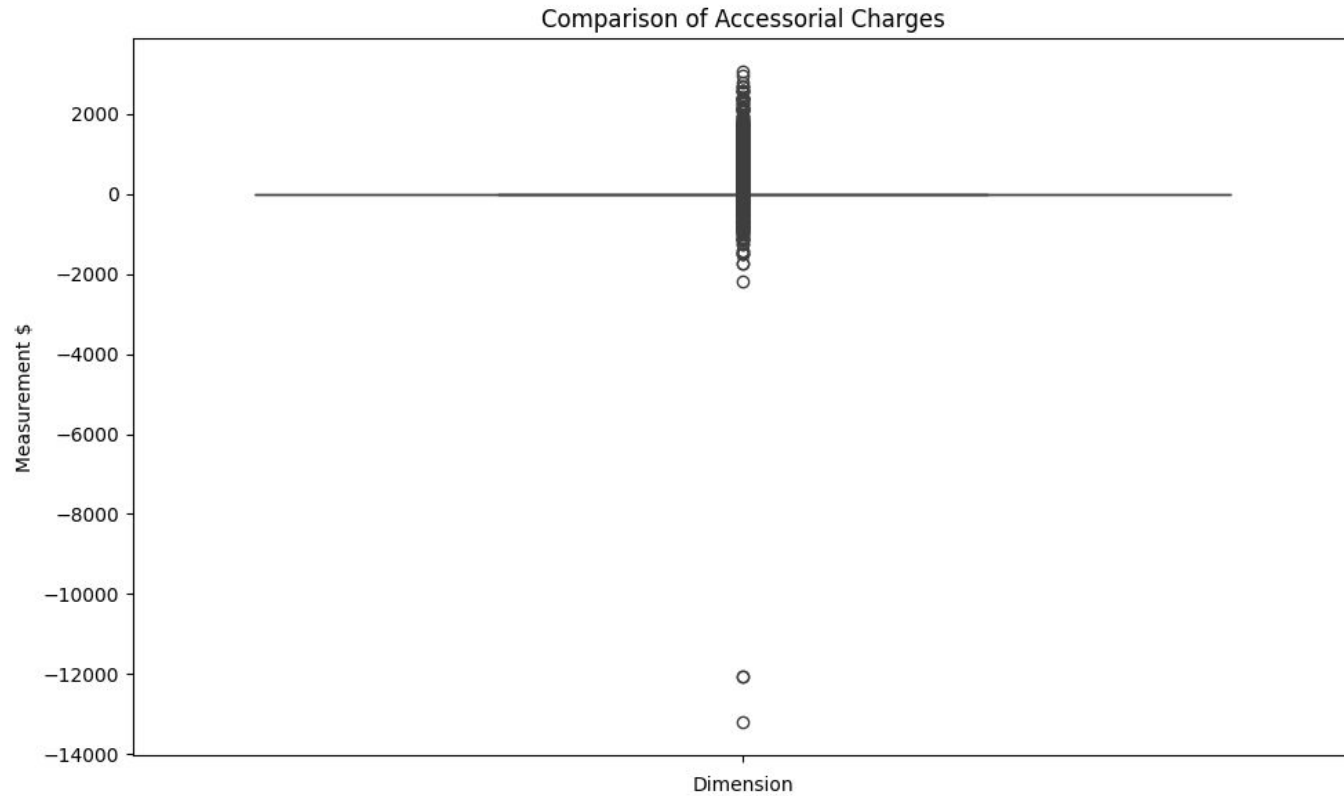


# Parcel

Number of Items Shipped by Top 10 Destinations



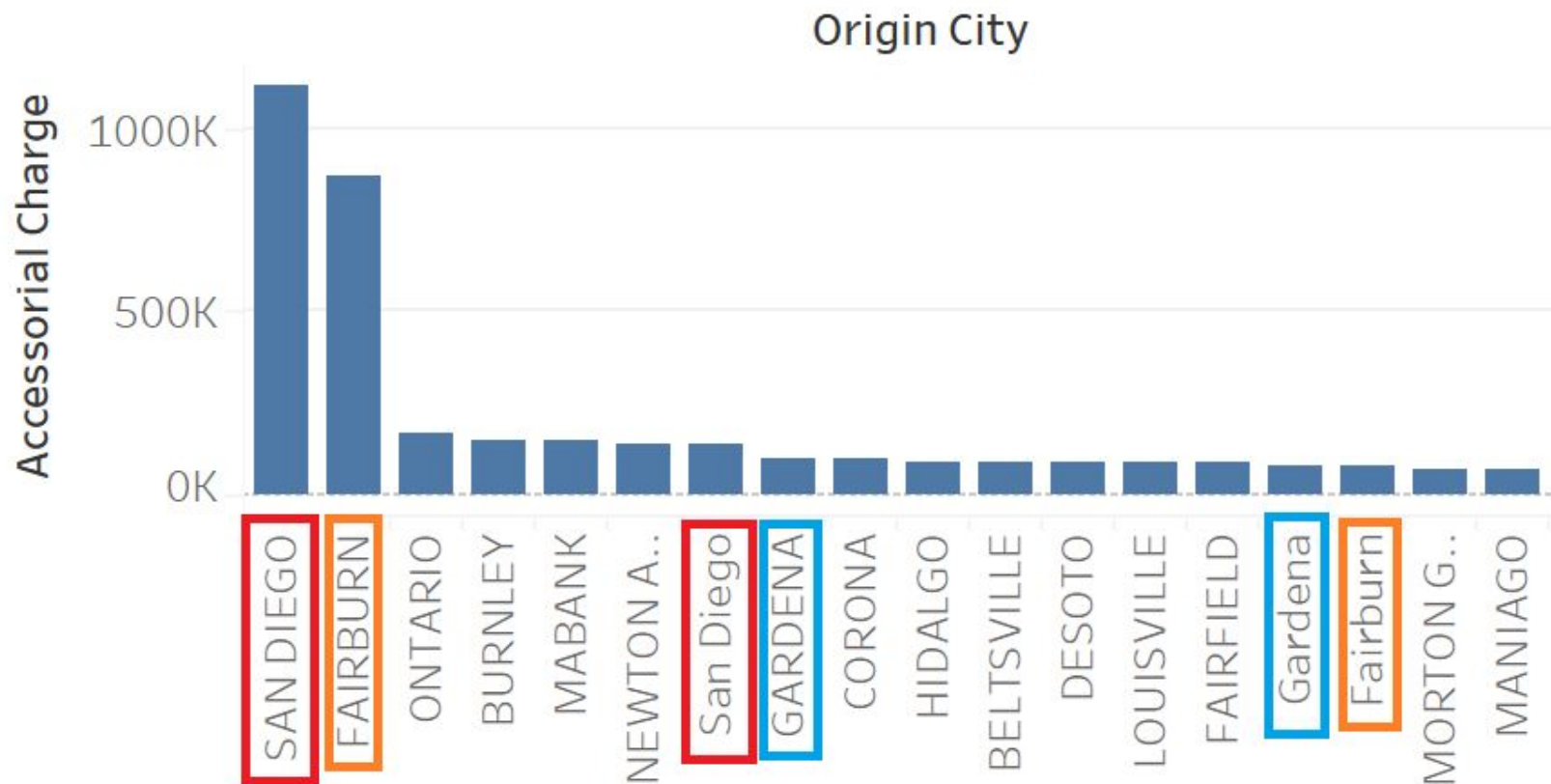
# Parcel



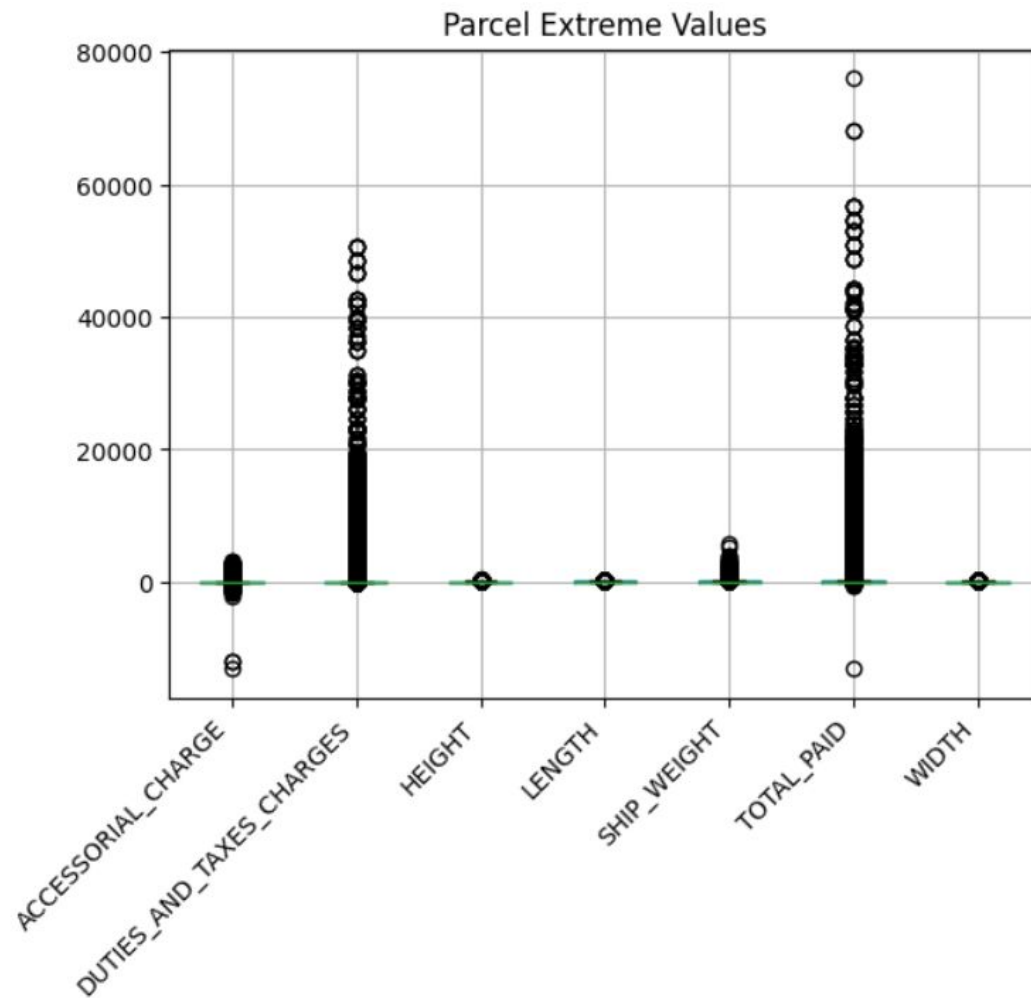
# Accessorial by Destination City



# Accessorial Charge By Origin City



# Parcel



Accessorial

# Accessorial Variables

Unnamed:	0
ACCESSORIAL_AMOUNT	0
ACCESSORIAL_CHARGE_AMOUNT	0
ACCESSORIAL_CHARGE_CODE	0
ACCESSORIAL_CHARGE_DESCRIPTION	0
AMOUNT_PAID	0
BILL_DATE	0
BILL_TYPE	0
CARRIER_MODE_DESCRIPTION	0
CARRIER_NAME2	0
DESTINATION_CITY	0
DESTINATION_COUNTRY_CODE	2813
DESTINATION_NAME2	0
DESTINATION_STATE	3286
DESTINATION_ZIP	153
IO_CODE	0

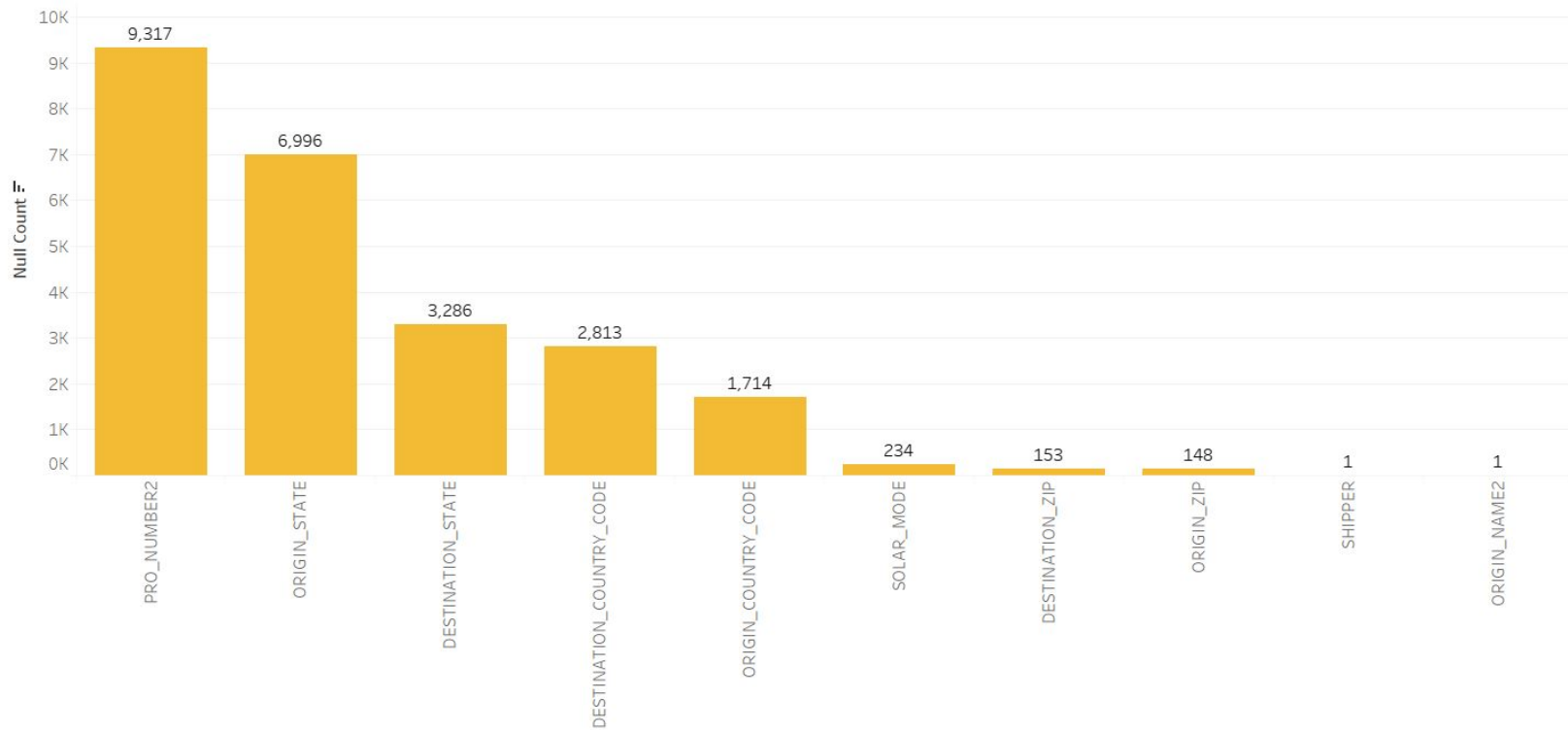
IO_CODE_DESCRIPTION	0
MILEAGE	0
ORIGIN_CITY	0
ORIGIN_COUNTRY_CODE	1714
ORIGIN_NAME2	1
ORIGIN_STATE	6996
ORIGIN_ZIP	148
PROCESS_DATE	0
PRO_NUMBER2	9317
RECEIVER	0
SHIPPER	1
SHIPPER_NAME	0
SHIP_DATE	0
SHIP_WEIGHT	0
SOLAR_MODE	234

Count: 278,587

# Accessorial

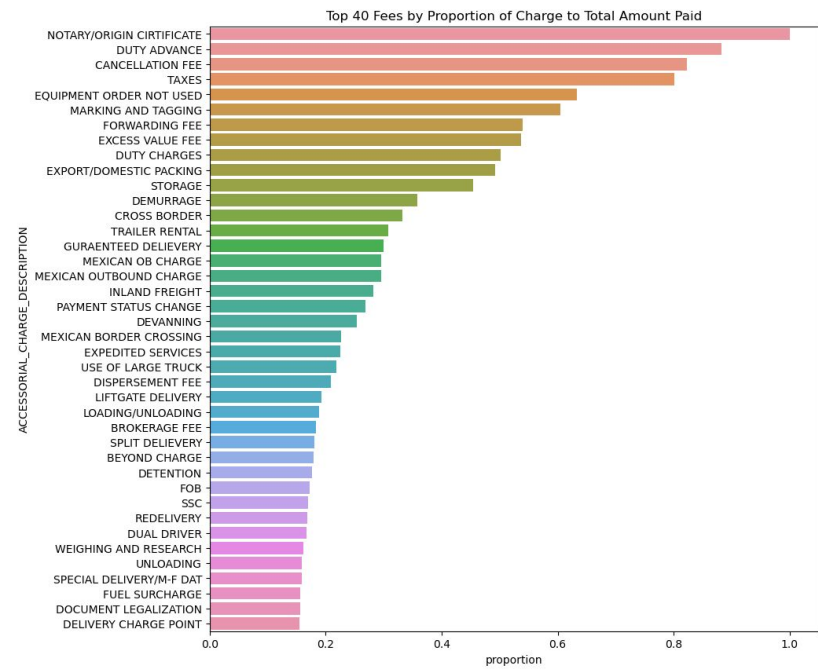
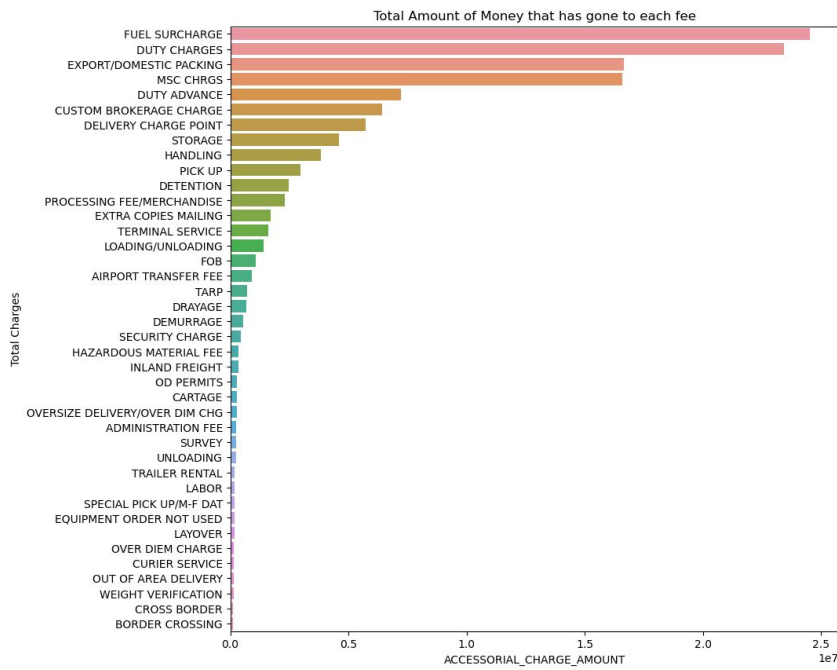
Null Count in *Accessorial*

n= 278,587

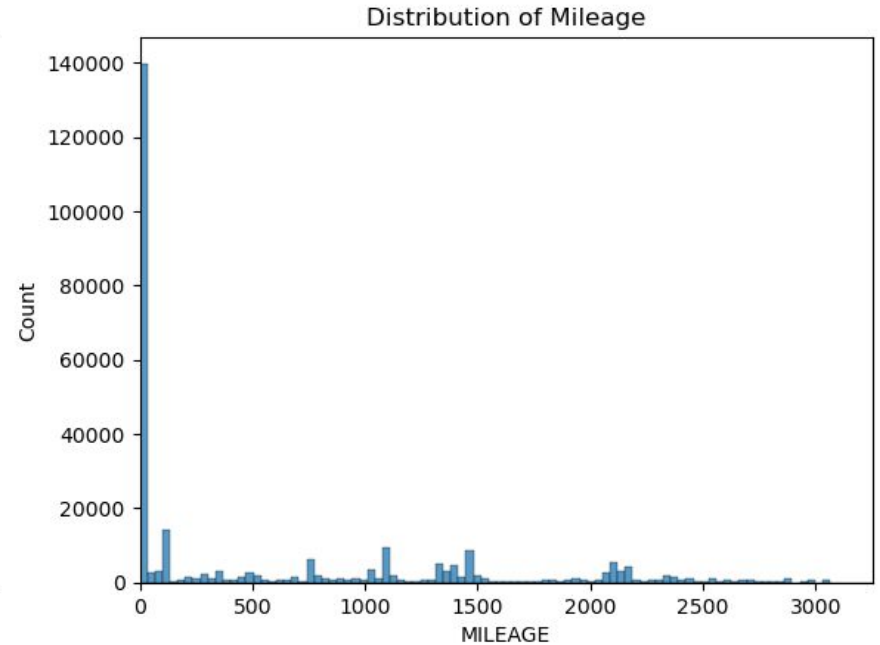
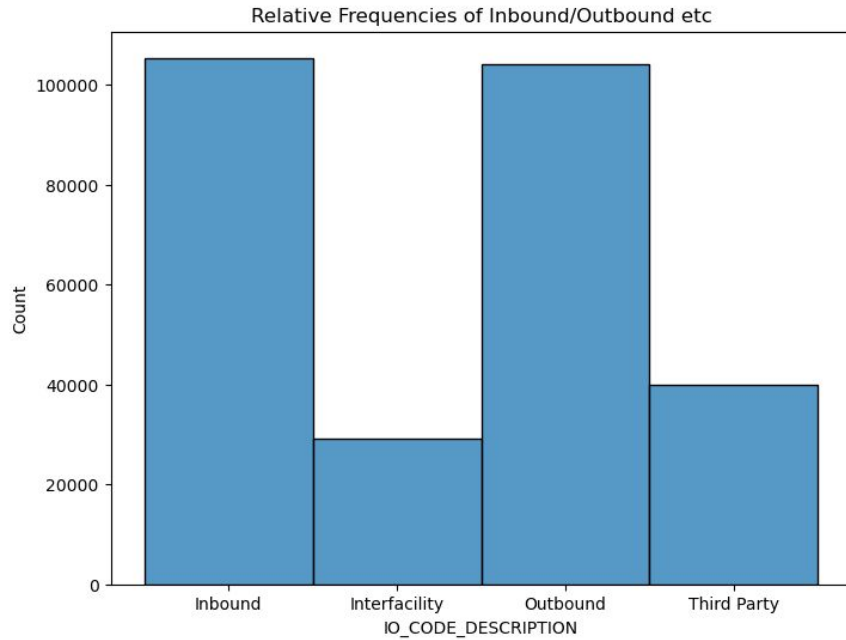




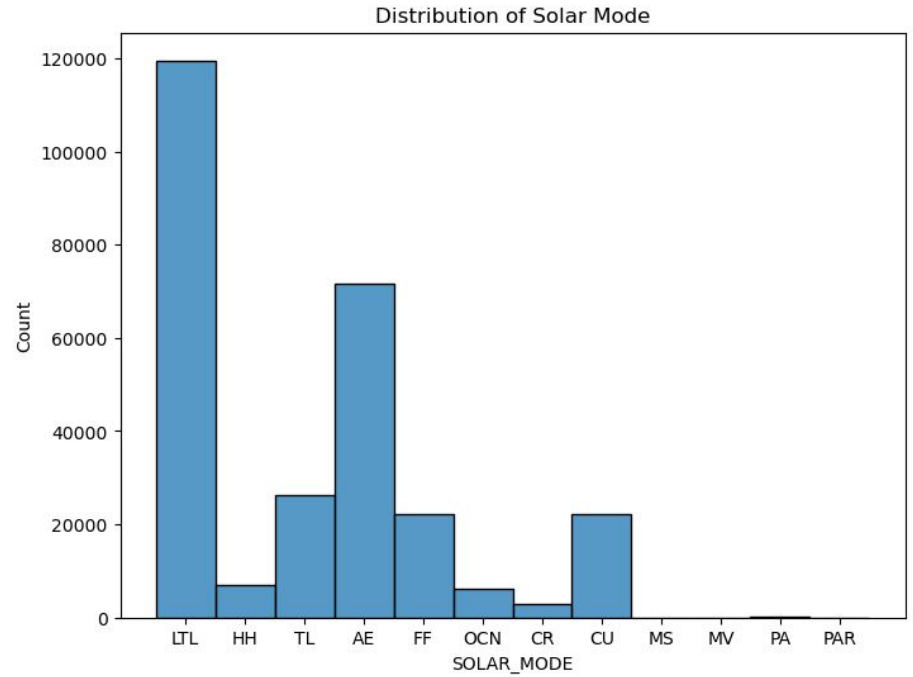
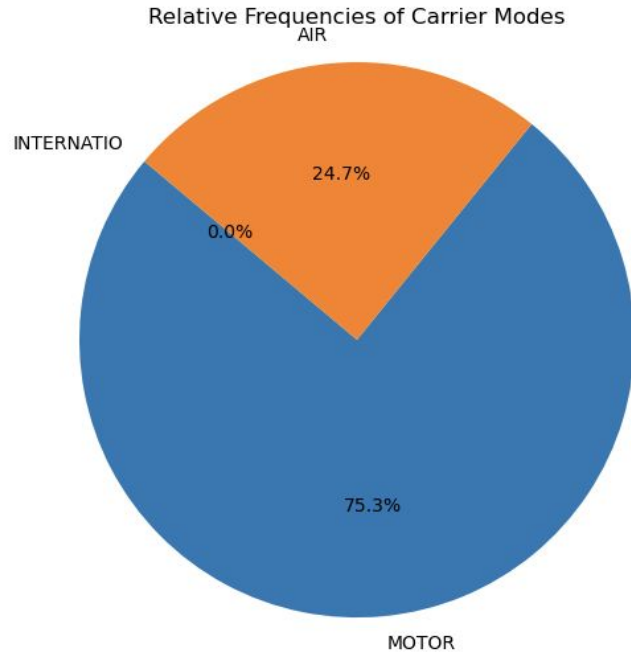
# Accessorial



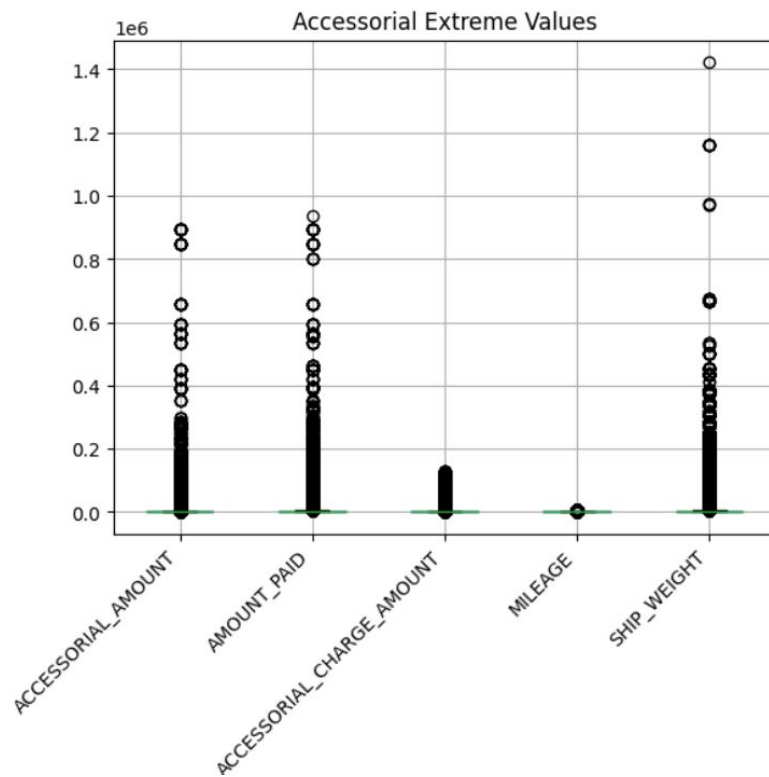
# Accessorial



# Accessorial



# Accessorial



EDI

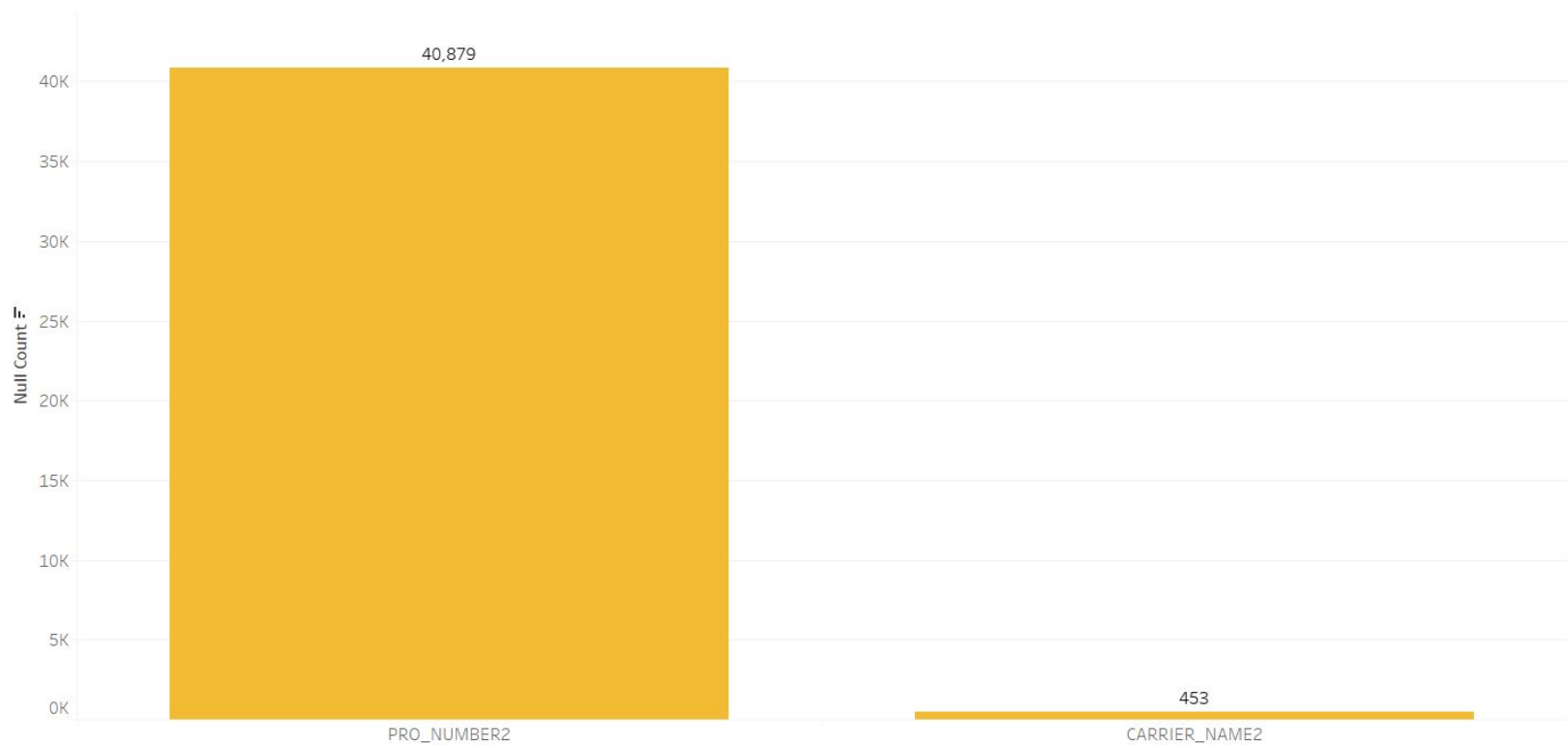
# ***EDI* Variables**

Unnamed:	0
AMOUNT_PAID	0
CARRIER_NAME2	453
EDI_PAPER_CODE	0
PRO_NUMBER2	40879

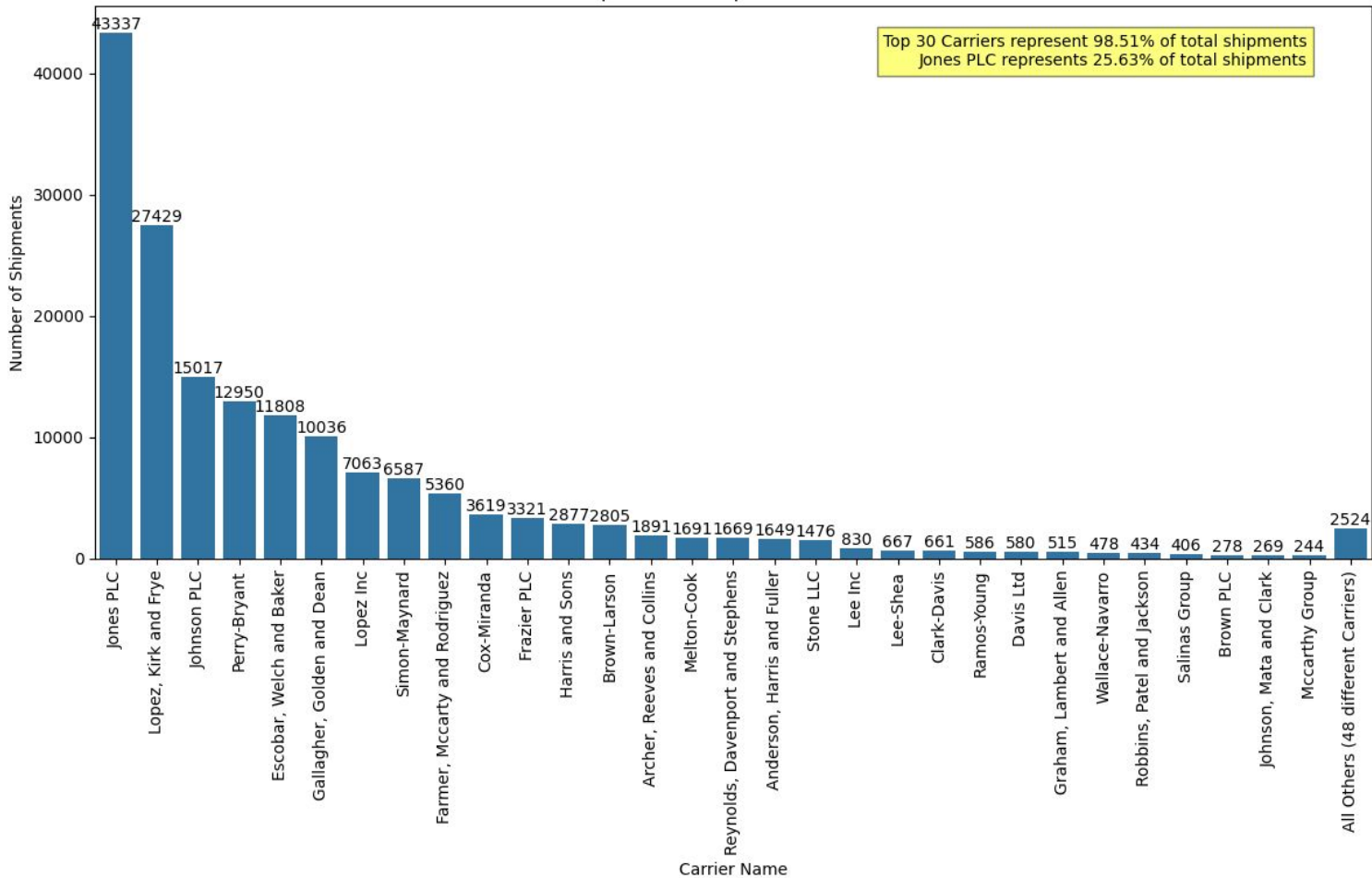
Count: 169,510

# EDI

Null Count in *EDI*  
n= 169,510

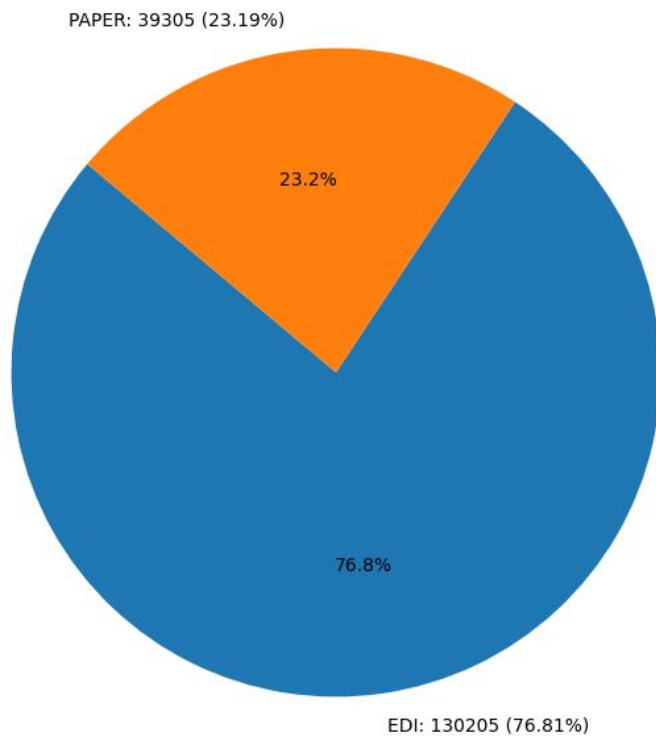


## Number of Shipments for Top 30 Carriers and All Others

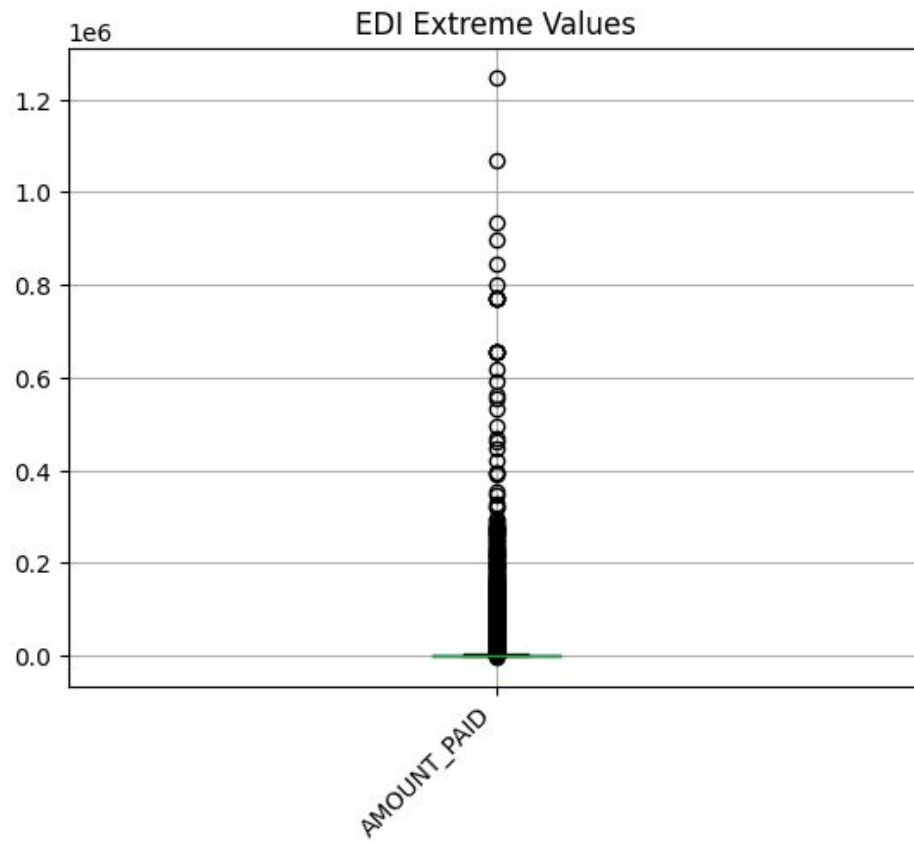




Distribution of EDI\_PAPER\_CODE



# EDI



Freight

# Freight Variables

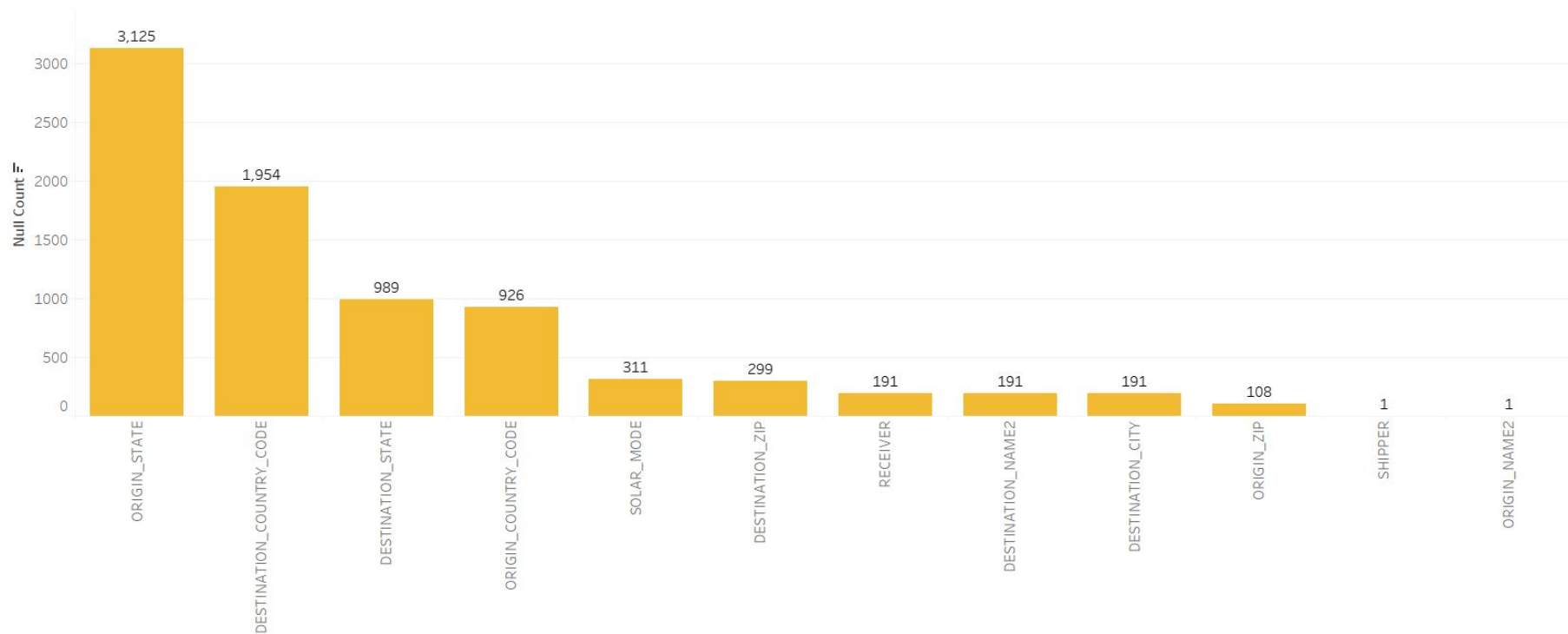
Unnamed:	0
ACCESSORIAL AMOUNT	0
AMOUNT PAID	0
CARRIER NAME2	0
DESTINATION CITY	191
DESTINATION CITY STATE	0
DESTINATION COUNTRY CODE	1954
DESTINATION NAME2	191
DESTINATION STATE	989
DESTINATION ZIP	299
IO CODE	0
MILEAGE	0
ORIGIN CITY	0
ORIGIN_CITY_STATE	0

ORIGIN COUNTRY CODE	926
ORIGIN NAME2	1
ORIGIN STATE	3125
ORIGIN ZIP	108
PROCESS DAY YEAR	0
PRO NUMBER2	0
RECEIVER	191
SHIPPER	1
SHIP DATE	0
SHIP DAY YEAR	0
SHIP WEIGHT	0
SOLAR_MODE	311

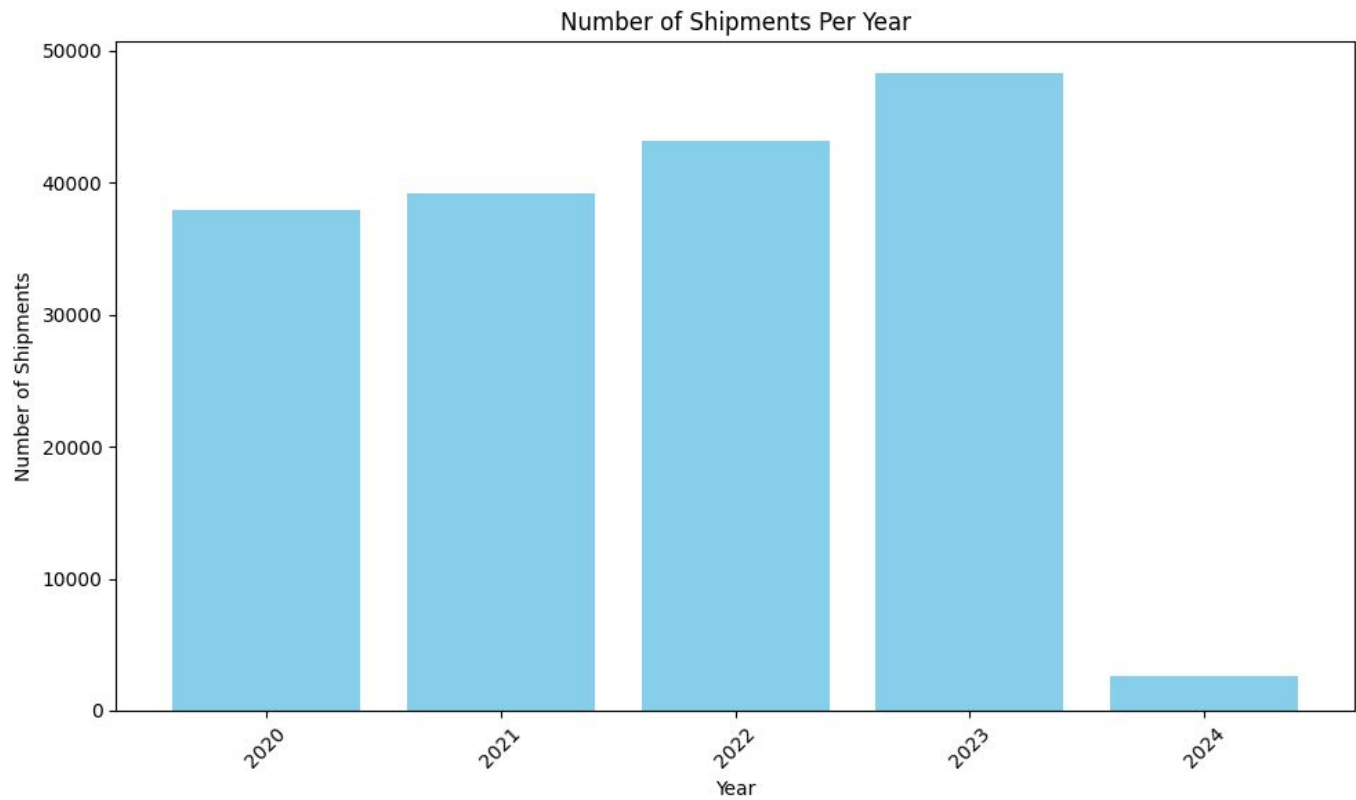
Count: 171,150

# Freight

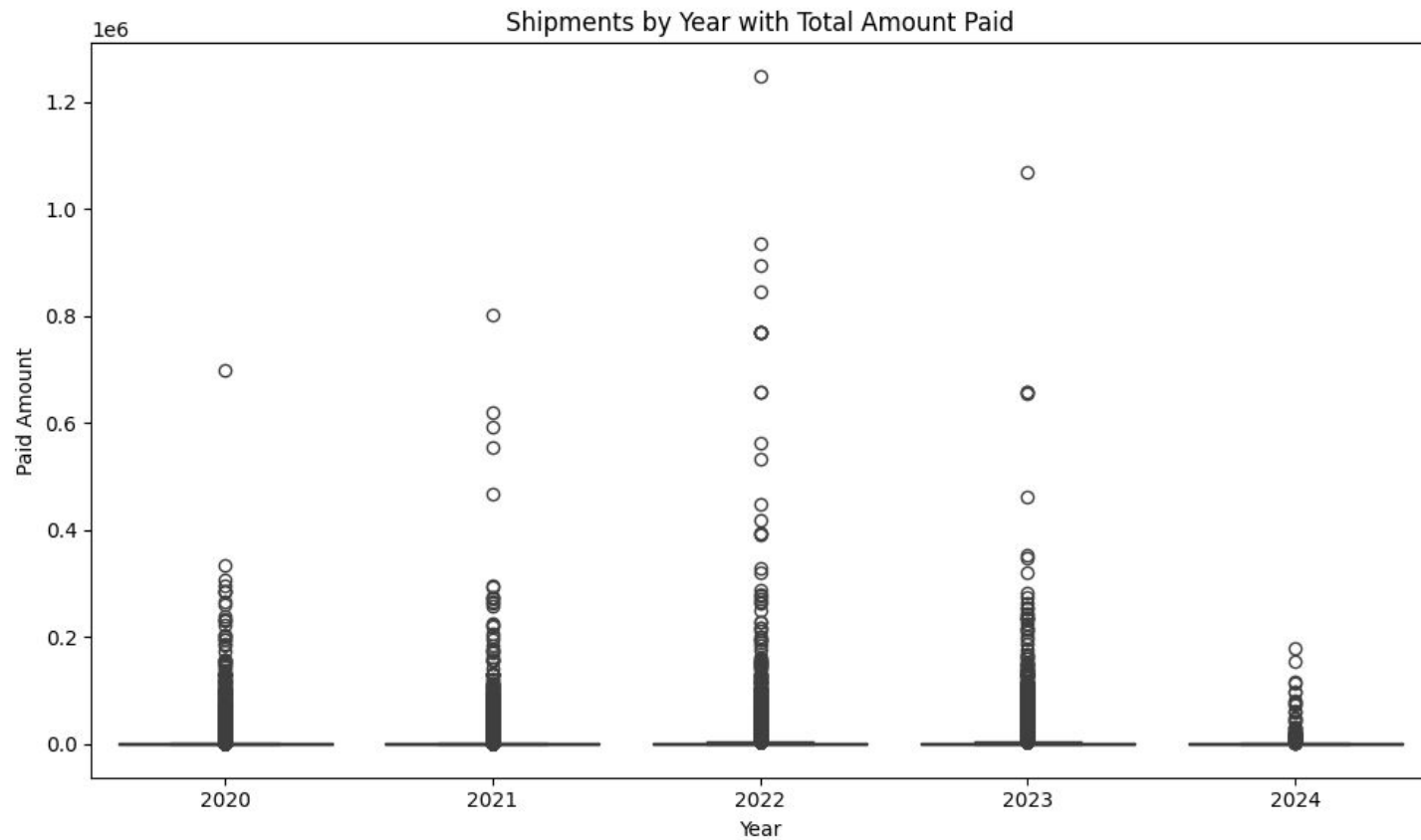
Null Count in *Freight*  
n= 171,150



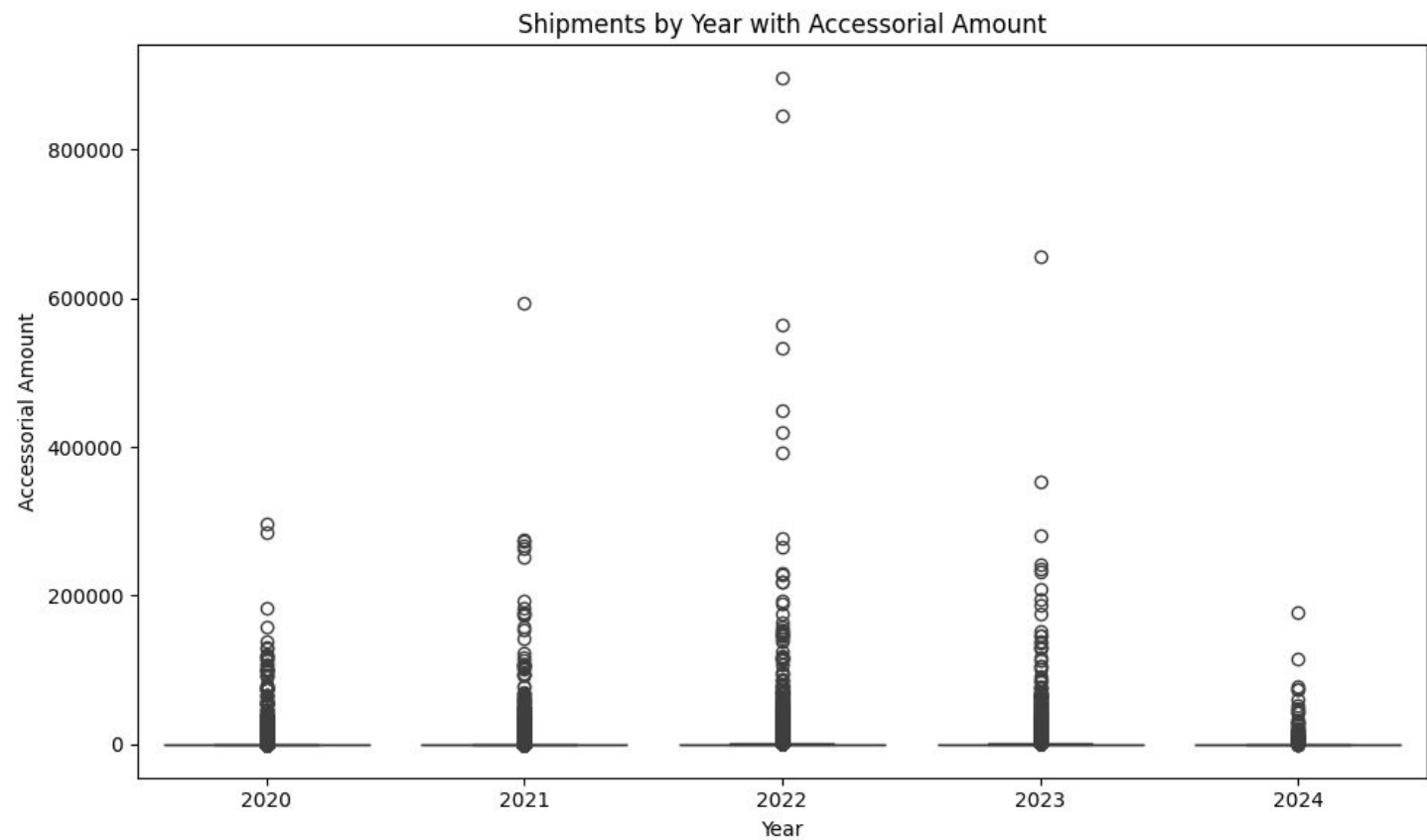
# Freight



## ***Freight***

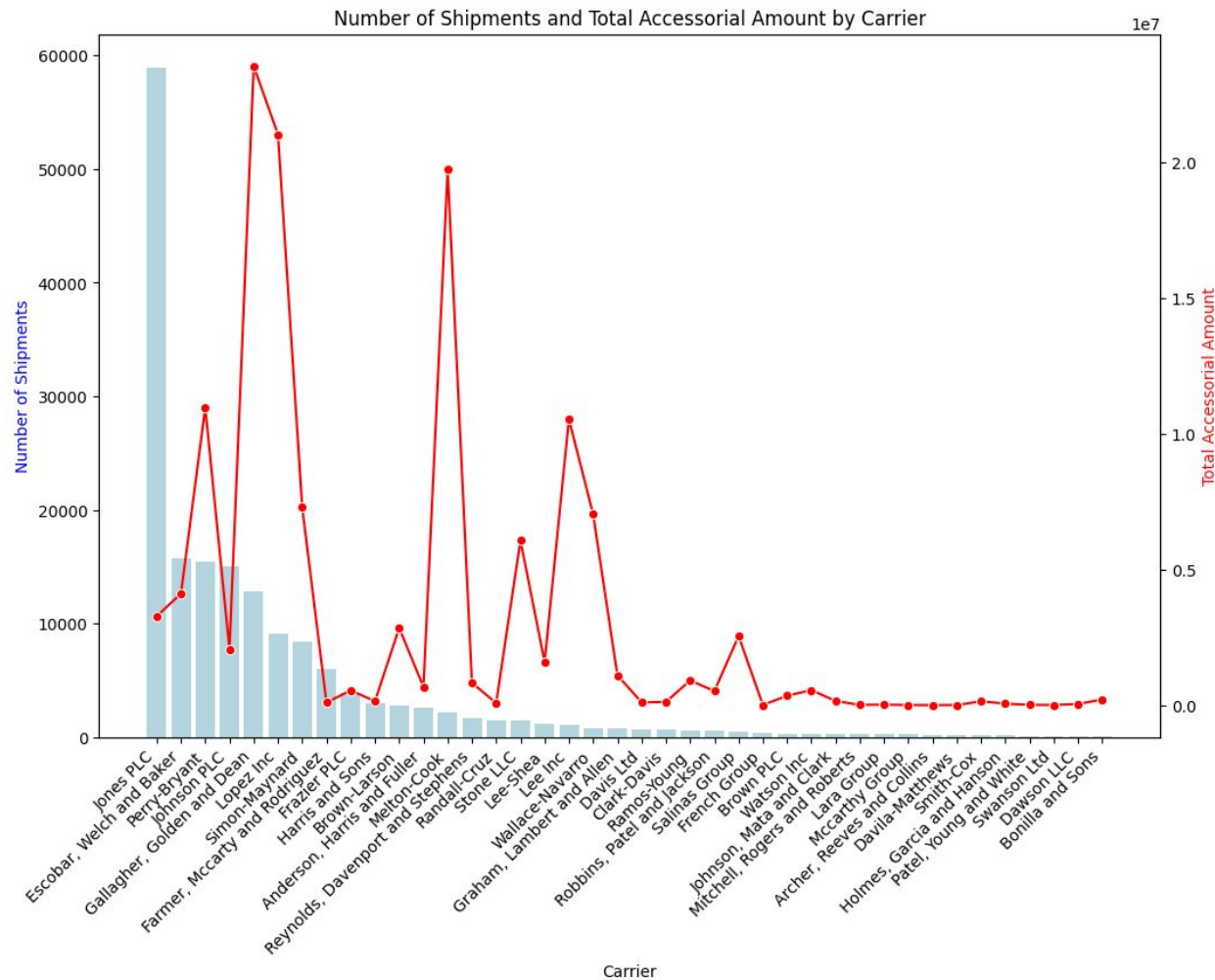


# Freight

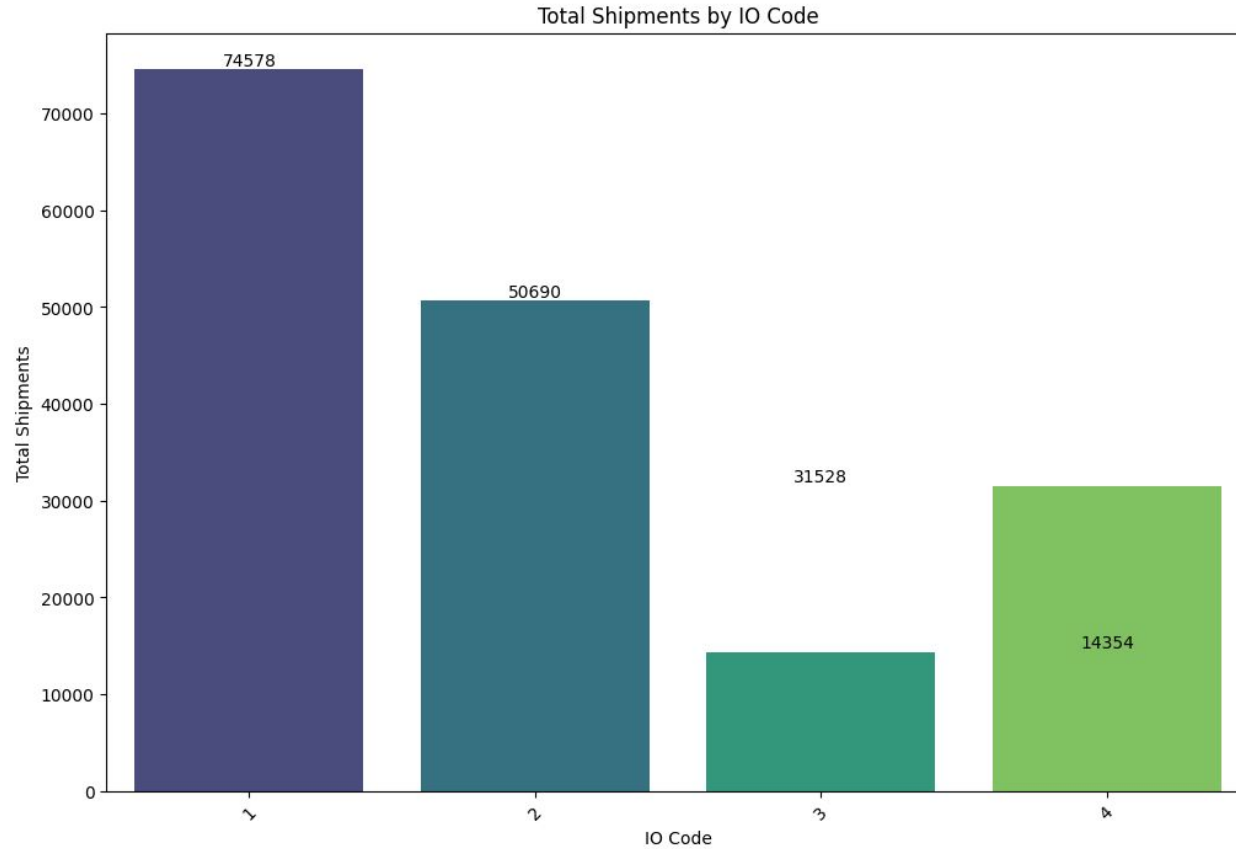




# Freight

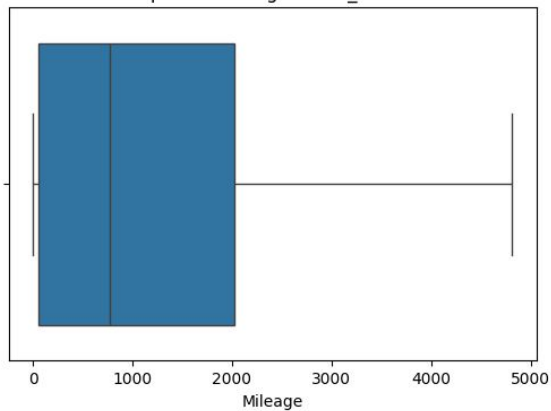


# Freight

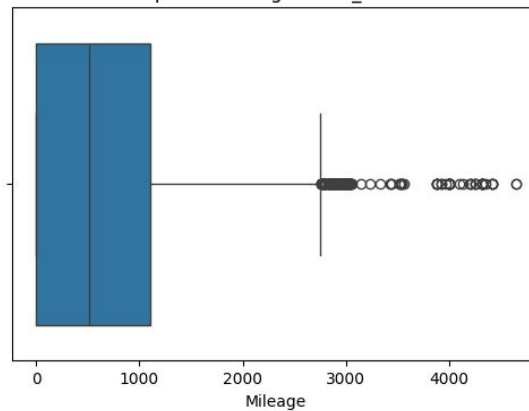


# Freight

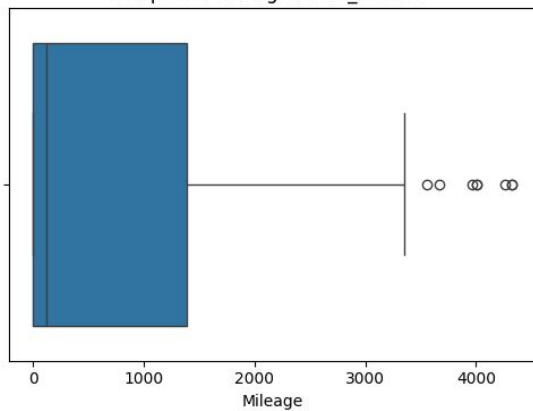
Boxplot of Mileage for IO\_CODE 1



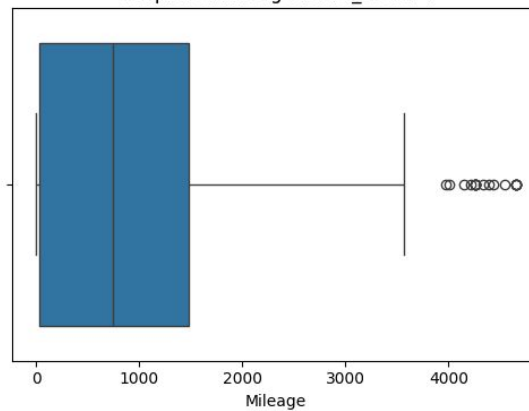
Boxplot of Mileage for IO\_CODE 2



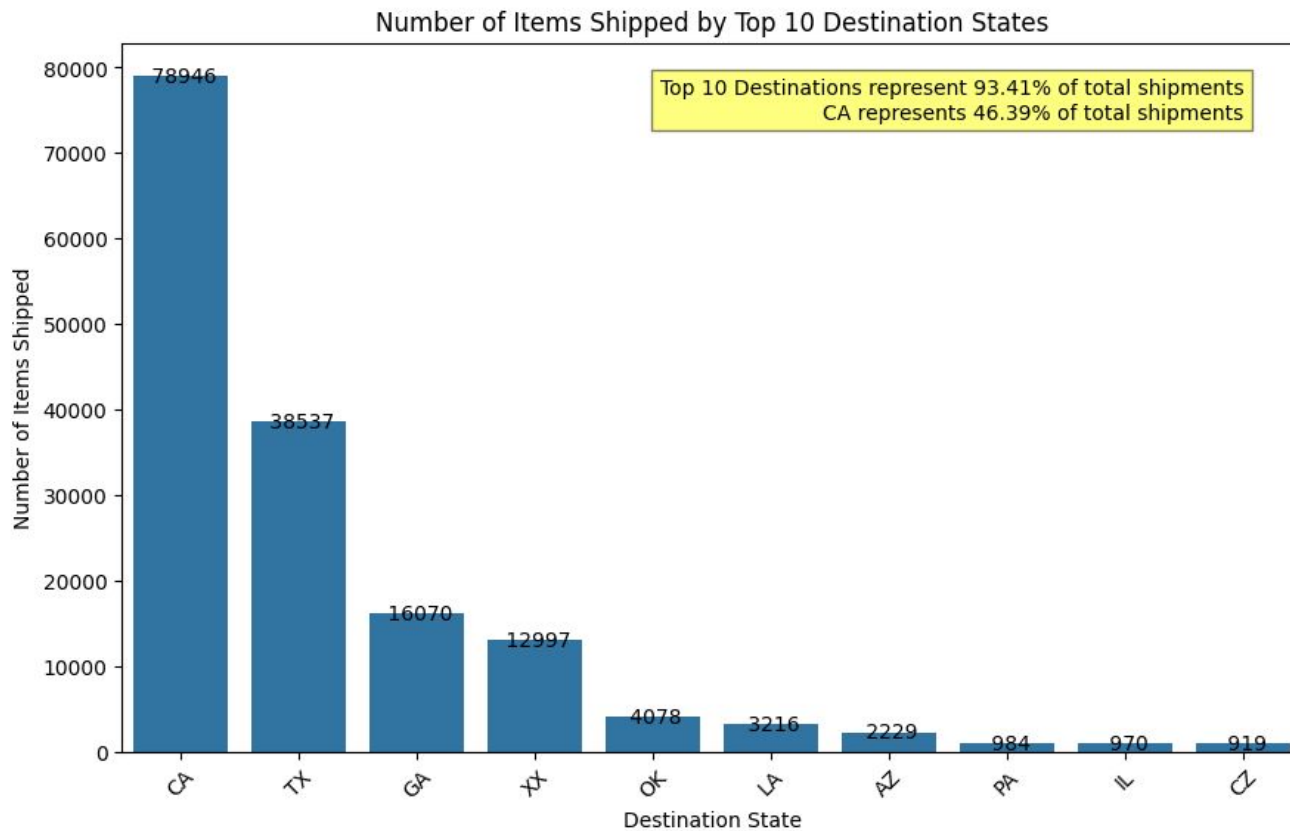
Boxplot of Mileage for IO\_CODE 3



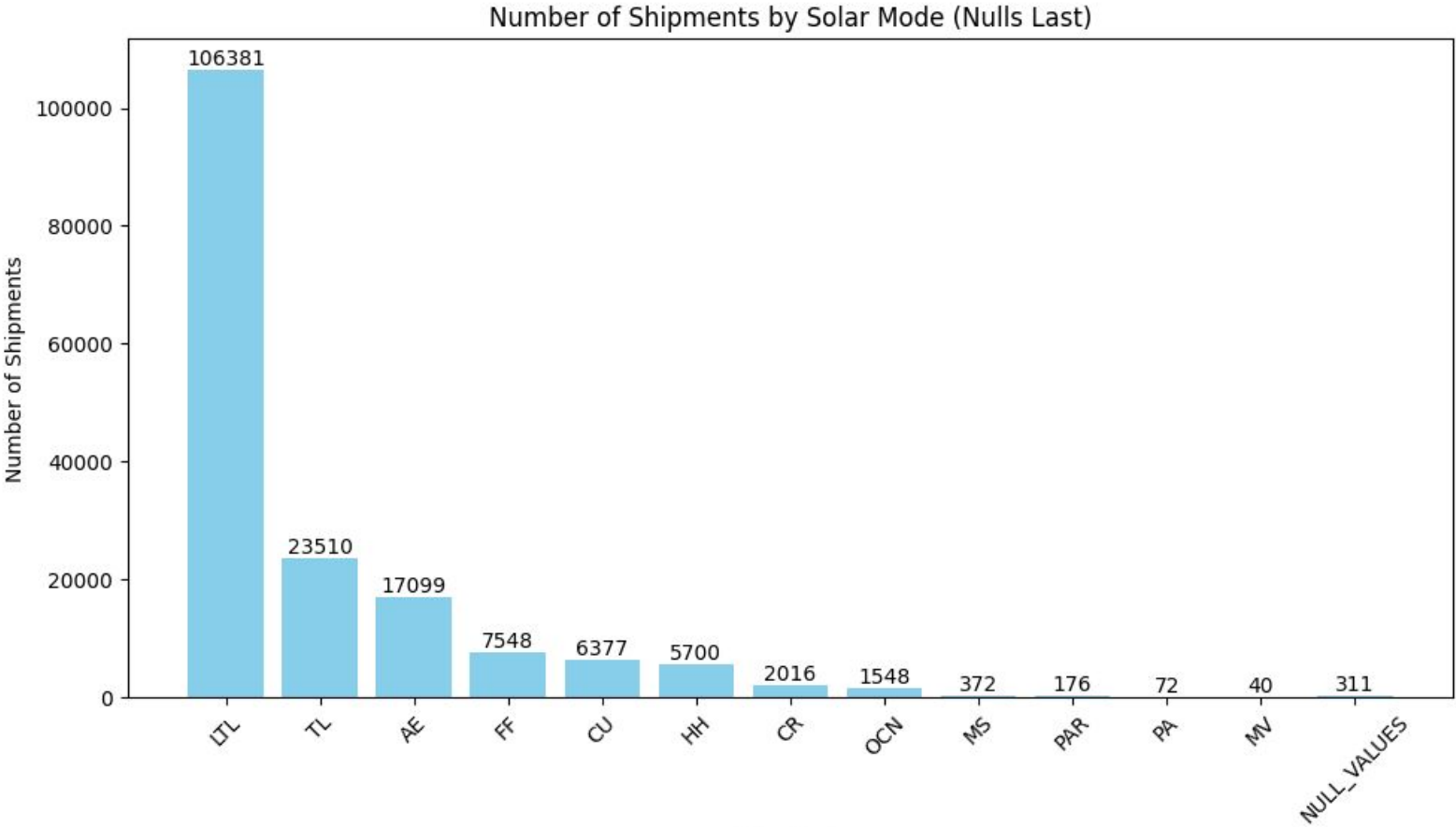
Boxplot of Mileage for IO\_CODE 4



# Freight



# Freight



# Freight

