

# Xpressbees Data Analyst Assignment

By : VAIBHAV AMRUT PALKAR

# Business Understanding

## Objective:

Exploratory Data Analysis of Logistics data for an organization to get insight and to improve the business process.

## Strategy:

By using R language, Tableau skills it need to find out observation and provide suggestions, Recommendations and Future scope aligning your objectives defined.

# Data Understanding

Here we have data for shipment flow from origin to destination location.

Data frame with total **25000 observations** and **29 variables**

We have different variable's as below

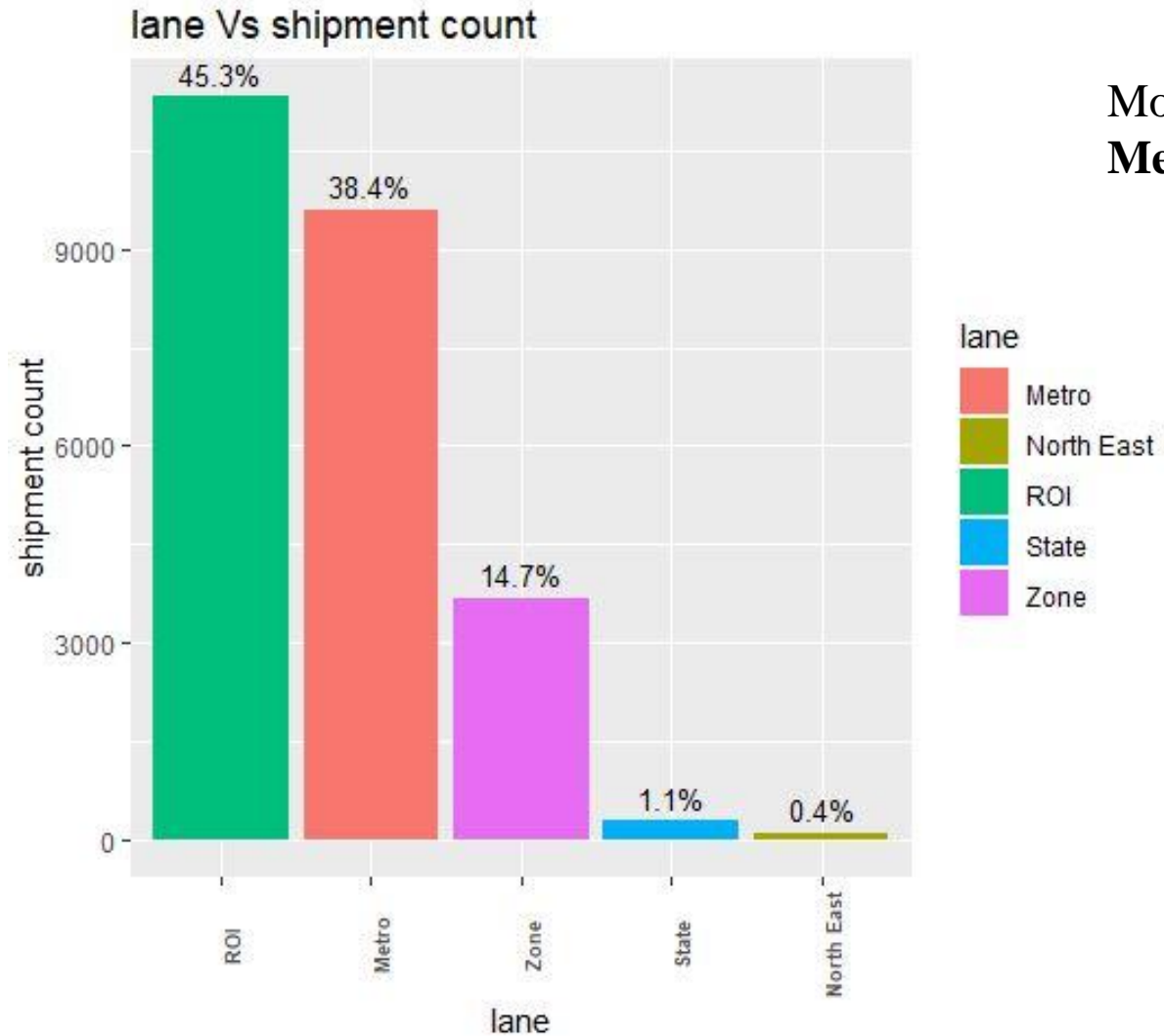
- Lanes - Shipment travelling in different Lanes Metro, North East, ROI, State, Zone
- Shipping ID - Unique ID for shipment
- Hubs - Unique ID for origin/destination
- Hub City/Hub Zone - City & Zone of Hubs
- Process Location - Touch Points of the shipment in its movement
- Process Location / Origin/Destination Baginscandate - Shipment scanned for date time once reaches the Touch Point/Origin/Destination
- Process Location/ Origin/Destination \_Bagoutscandate - Shipment scanned for date time once it moves out from Touch Point/Origin/Destination
- Delivery Date - shipment delivery date time
- Physical wt/vol wt - physical wt in kg/vol wt in mtr.cube
- Ship Pincode - Destination Pincode
- Route Mode - Mode in which shipment has travelled

# Data Manipulation

- Logistics shipping data file
- Data cleaning checks for erroneous data:
  - NA values – 0 NA values found
  - Blank values – 0 blank values found
  - NULL values – total 259 NULL values found (ProcessLocation3\_BagOutScanDate – 7 and Destination BagOutScanDate - 252)
  - Duplicate record – 0 duplicates record for unique shipping id column
  - Treated route mode AIR and Air to single value Air
- Feature Engineering :
  - Date extraction and conversion to standard date format using Lubridate package
  - Created new variables of time for shipments and shipment processing
  - Created new variable and data frames for analysis purpose
  - Created cost variable by assuming shipment cost
  - Created data frame for NULL values for separate analysis

# Univariate Data Analysis

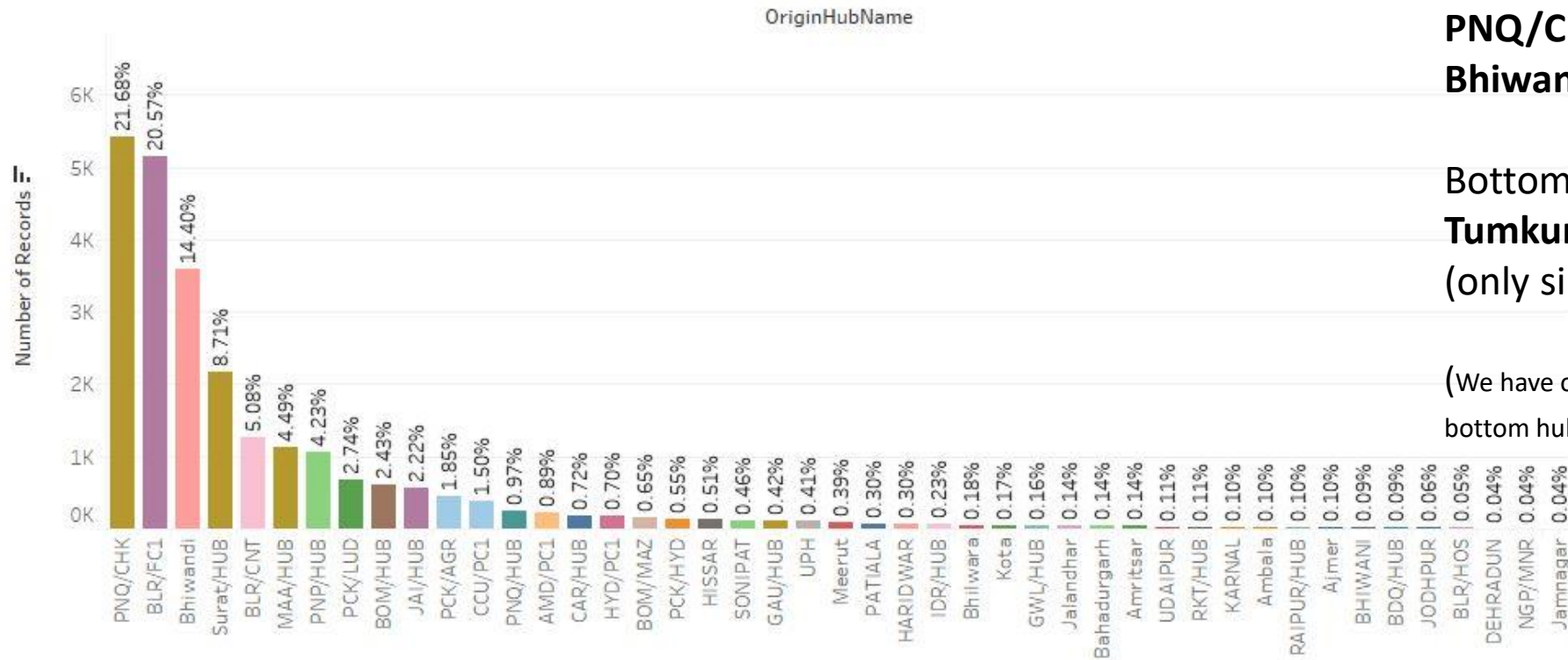
# Lane



Most of the shipments are happened between **ROI** (45.3%) and **Metro** (38.4%) and only 16.2 % between other lanes.

# Origin Hub Name

Origin Hub Name vs Shipment Count



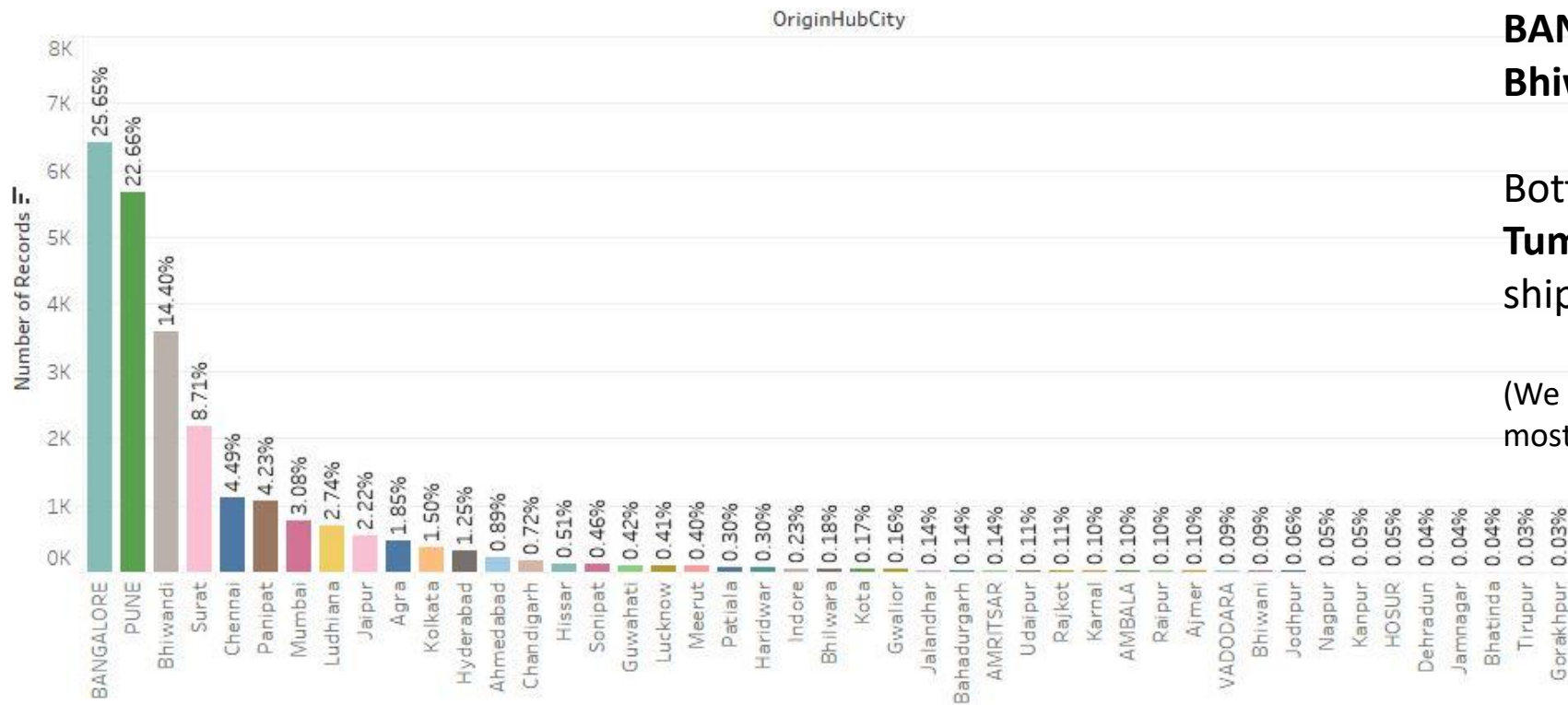
Top 3 processing origin Hubs - **PNQ/CHK** – 21.7%, **BLR/FC1** – 20.6\$, **Bhiwandi** – 14.4%.

Bottom 3 origin Hubs shipment - **Tumkur, Udhamapur , VPI/HUB** (only single shipment).

(We have only top 45 Hubs in visualization as most bottom hubs have few or 1 records)

# Origin Hub City

Origin Hub City vs Shipment Count



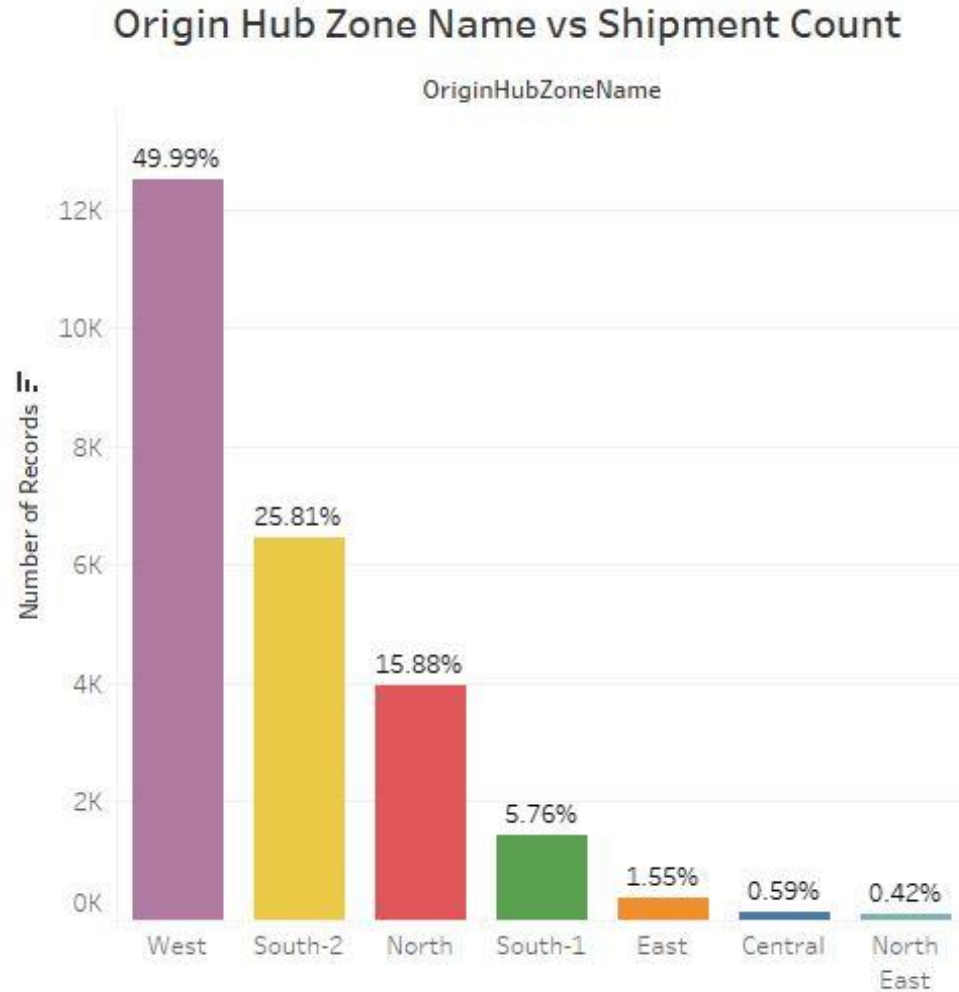
Top 3 origin Hubs shipment -  
**BANGALORE** – 25.7%, **Pune**– 22.7%,  
**Bhiwandi** – 14.4%.

Bottom 3 origin Hubs shipment -  
**Tumkur, Udhampur , VPI** (only single shipment).

(We have only top 45 Hubs in visualization as most bottom hubs have few or 1 records)



# Origin Hub Zone Name



Most of the shipments are happened from **West** (50.0%) and **South 2** (25.8%) total south (25.8+5.8=31.6%).

And only 19.4 % between other zones.

# Ship Pin Code

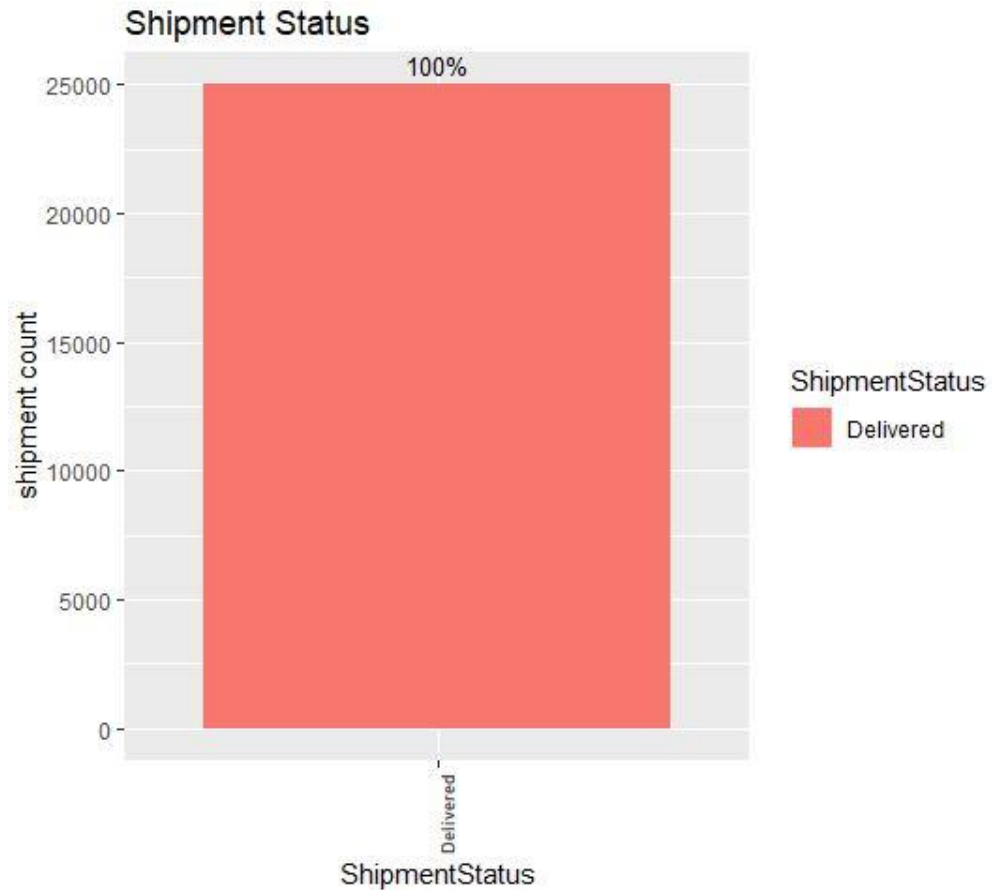
Most of the shipments going to pin code **201301** (7.10%) and **122001** (6.0%), which going to cities Noida, Delhi NCR, Gurugram and Gurgaon.



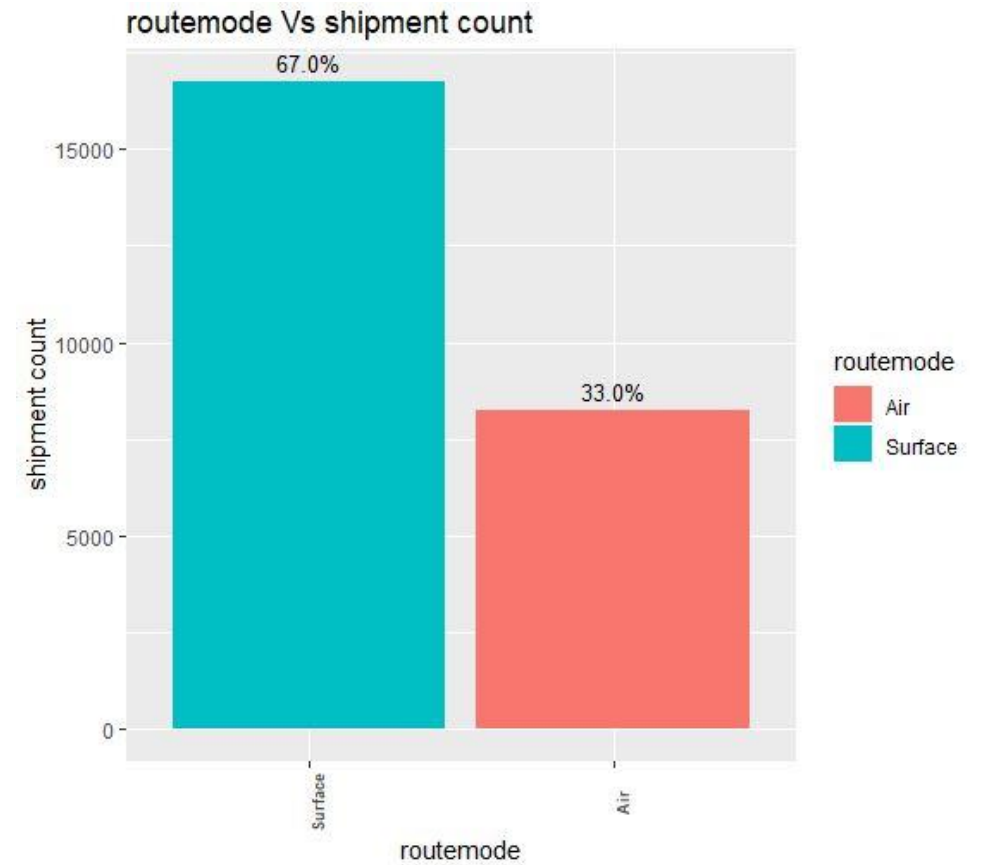
Then there 17 ship pin code where less than 10 shipments are going per pin code.

# Shipment Status and Route Mode

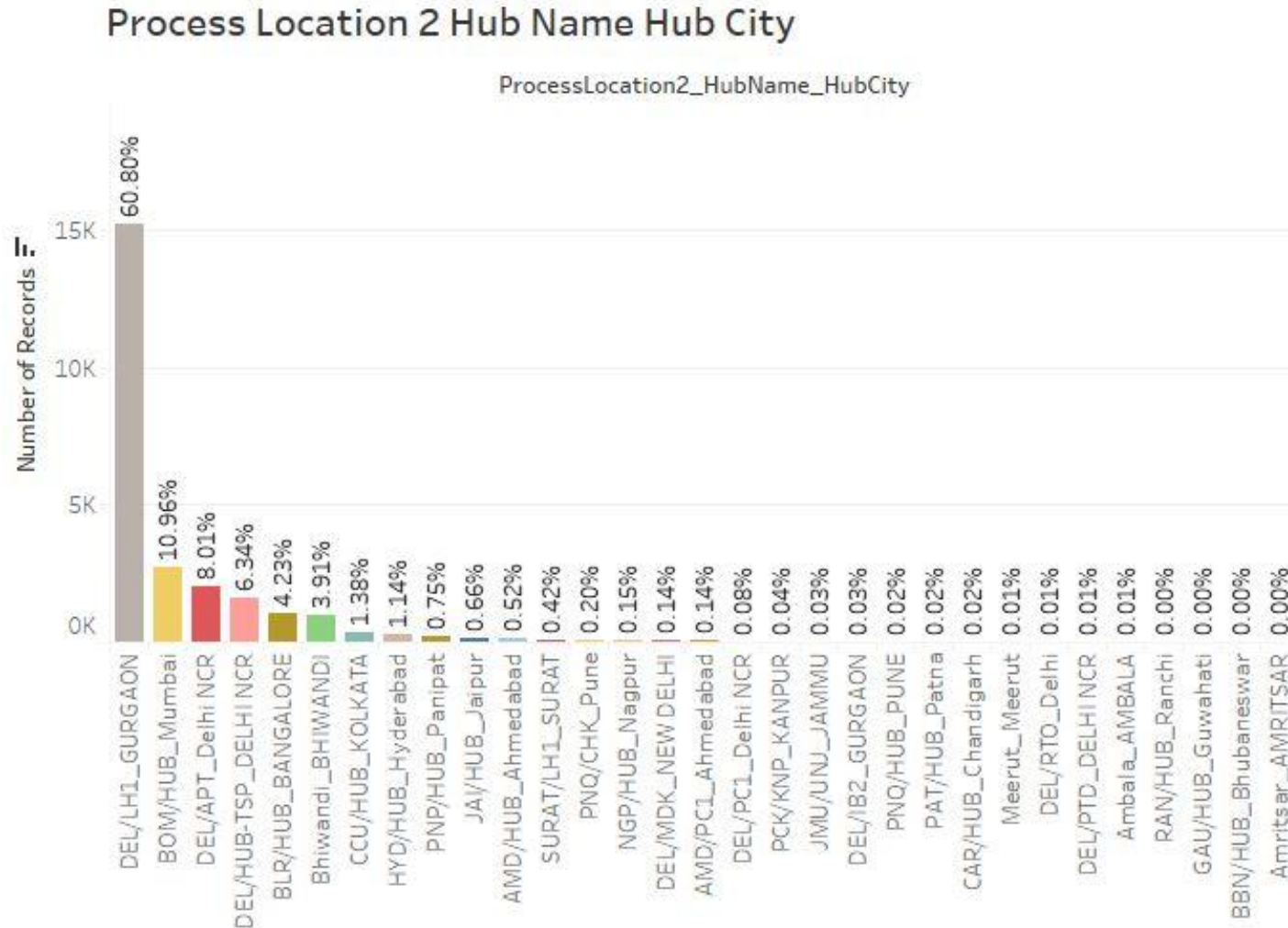
All shipments are delivered.



Most of the shipments are happening by **Surface mode** (67%) and other by **Air** (33%).



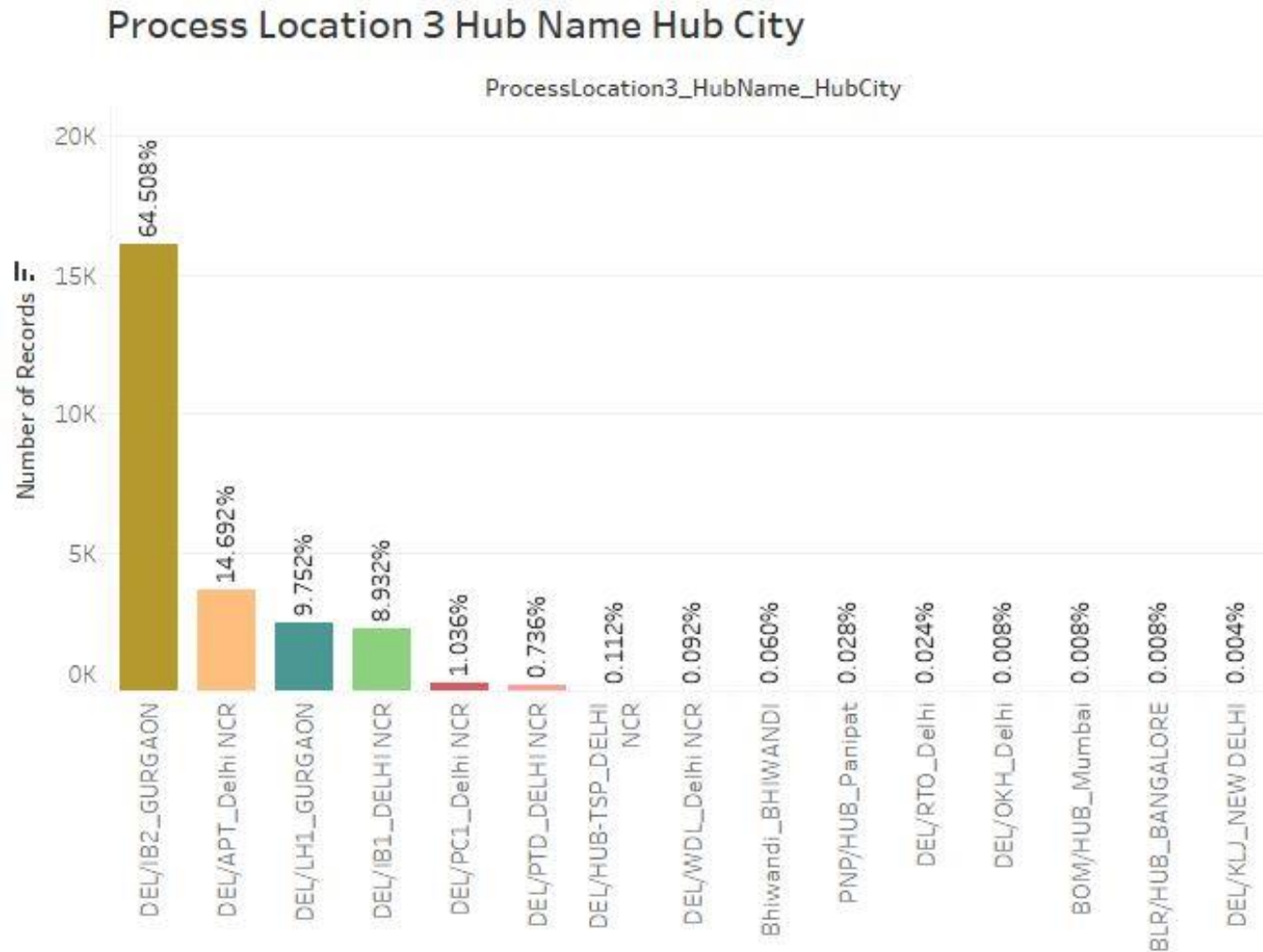
# Process Location 2 Hub Name and Hub City



Most of the shipments are process at **DEL/LH1\_Gurgaon** (60.8%) from which 83% shipped by surface route mode.

and **BOM/HUB\_Mumbai** (11%).

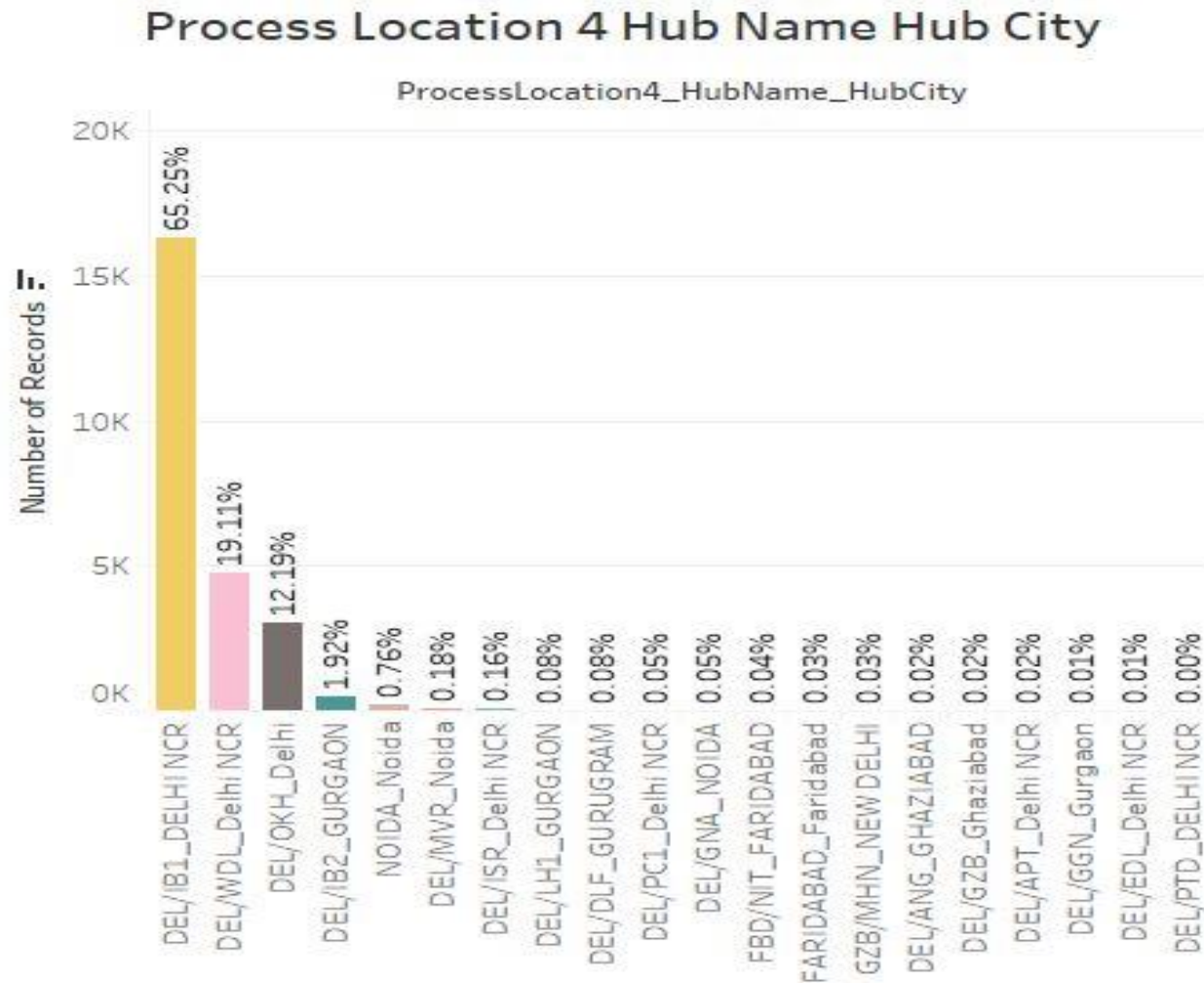
# Process Location 3 Hub Name and HubCity



Most of the shipments are process at **DEL/IB2\_Gurgaon** (64.5%) and we can see for most of other shipment processing location is Delhi city (because destination is in north).

And only 26 shipments(0.104%) are processed at other hub/cities Mumbai, Bhiwandi, Bangalore, Panipat.

# Process Location 4 Hub Name and HubCity

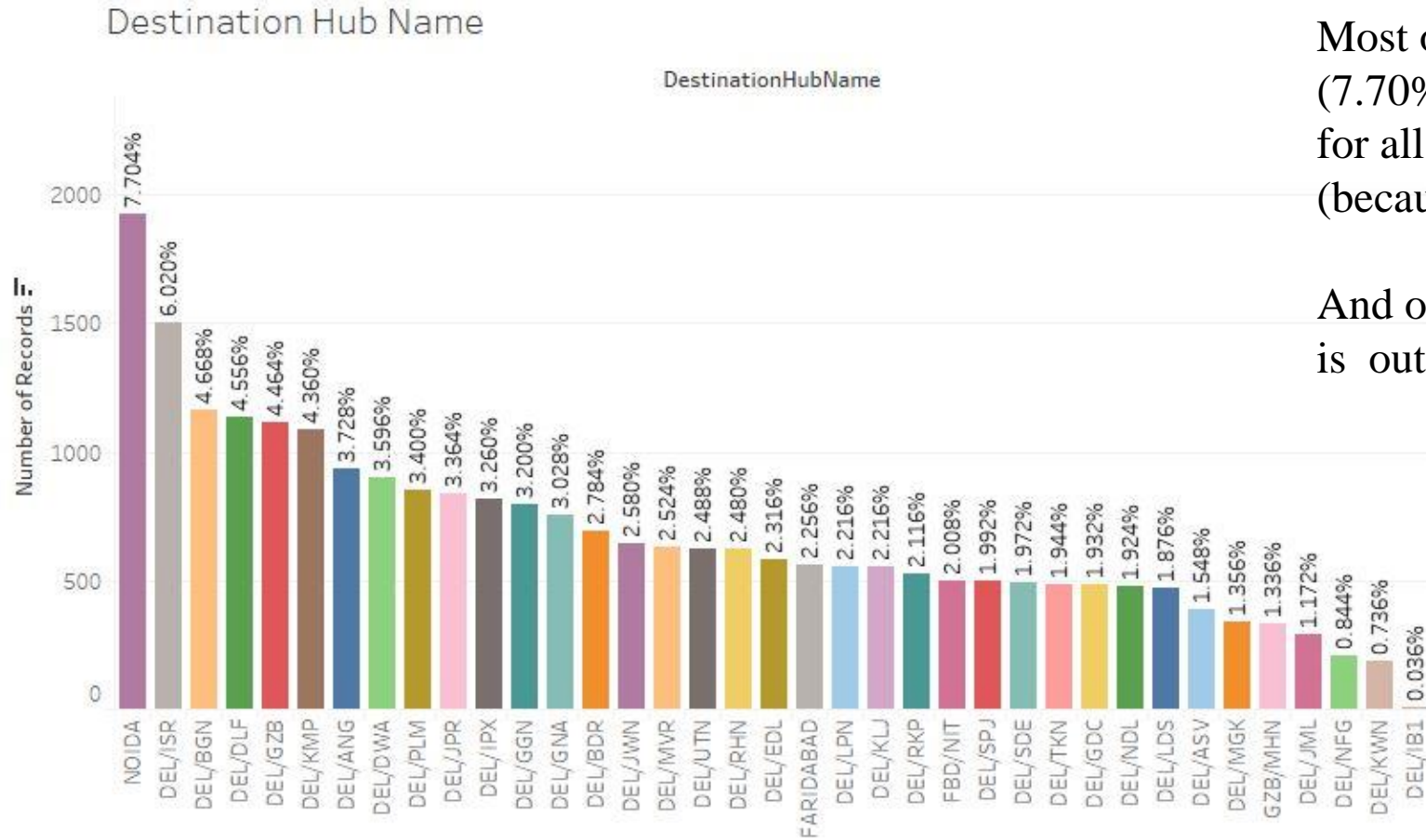


Most of the shipments are process at **DEL/IB1\_DELHINCR** (65.25%) and we can see for most of other shipment processing location is Delhi city (because destination is in Delhi in North Zone).

And only 216 shipments(0.864%) are processed at other hub/cities Mumbai, Bhiwandi, Bangalore, Panipat.



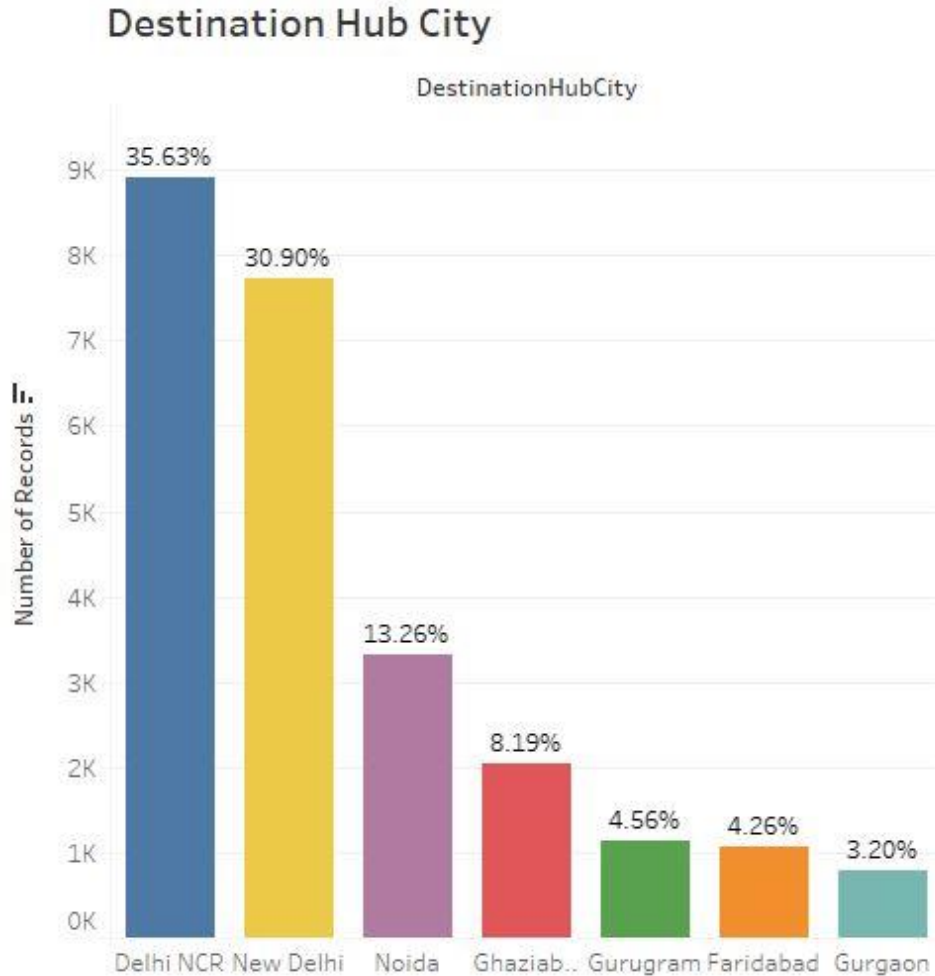
# Destination Hub Name



Most of the shipments destination hub at **Noida** (7.70%) and **DEL/ISR** (6.02%) and we can see for all shipment destination location is Delhi city (because destination is in Delhi in North Zone).

And only 3 hubs (13.31% shipments) destination is out of Delhi.

# Destination Hub City



Most of the shipments destination city is **Delhi NCR** (35.6%) and **New DELHI** (30.9%).

And only 33.5% shipments destination is out of Delhi (Noida, Ghaziabad, Gurugram).

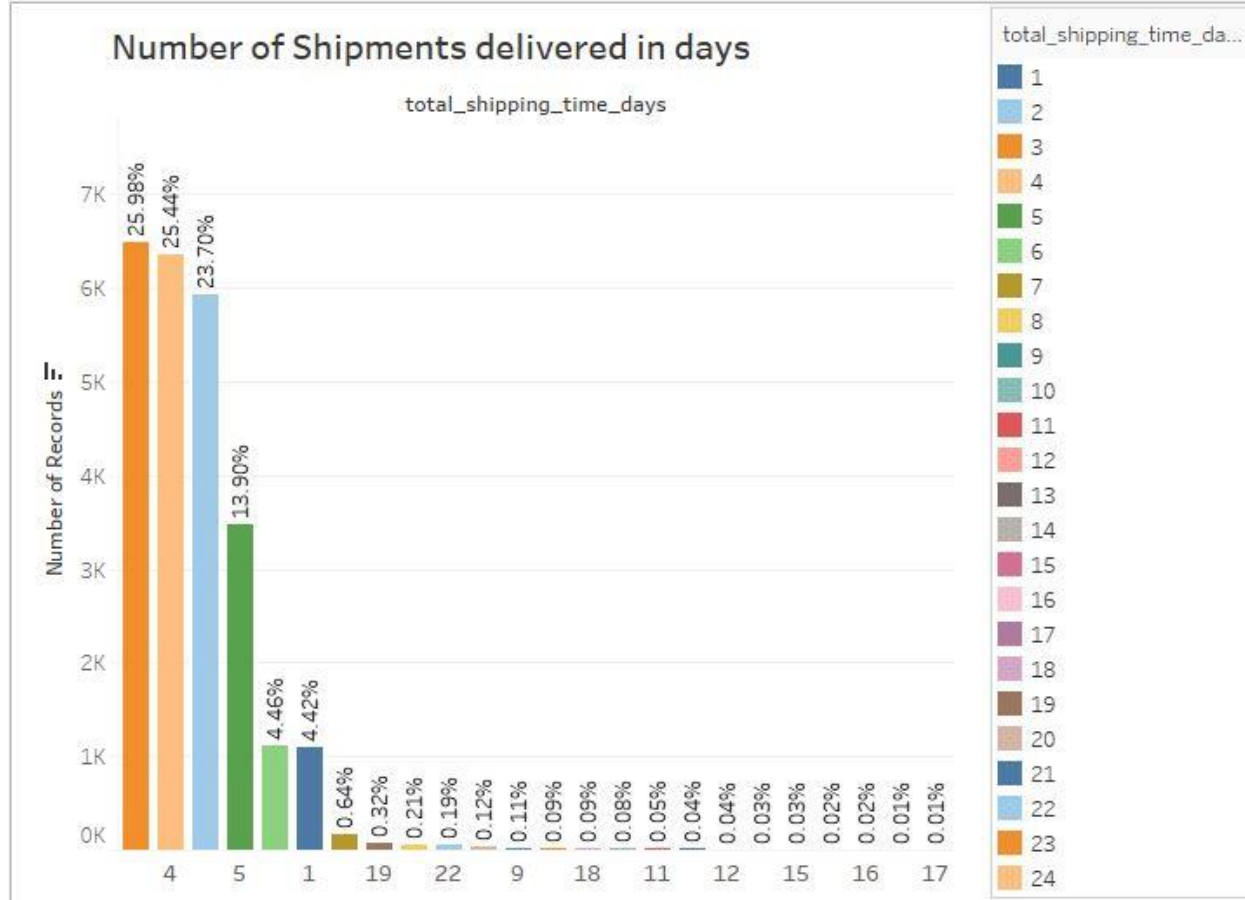


# Physical and Volumetric Weight

- Actual cost of the shipment is based on the physical or volumetric weight of the shipment.
- Physical weight – It is actual weight in kg of the shipment and is ranging from 0 to 34, and here 65% shipments have weight less than equal to 0.5 kg.
- Volumetric weight – It is volumetric weight of shipment in m<sup>3</sup> and is ranging from 0 to 74, and here 57% shipments have weight less than equal to 0.5 cubic meter.
- By assuming cost per kg(100 and 80 Rs) and per cubic meter(100 and 160 Rs), The charges for volumetric weight is more than by physical weight.
- Also shipment by air is more costlier than surface
- When physical and volumetric weights are zero then there should be fixed charges to be considered (as there are some shipments with 0 weight).

# Exploratory Data Analysis

# Shipment Delivery Time in days

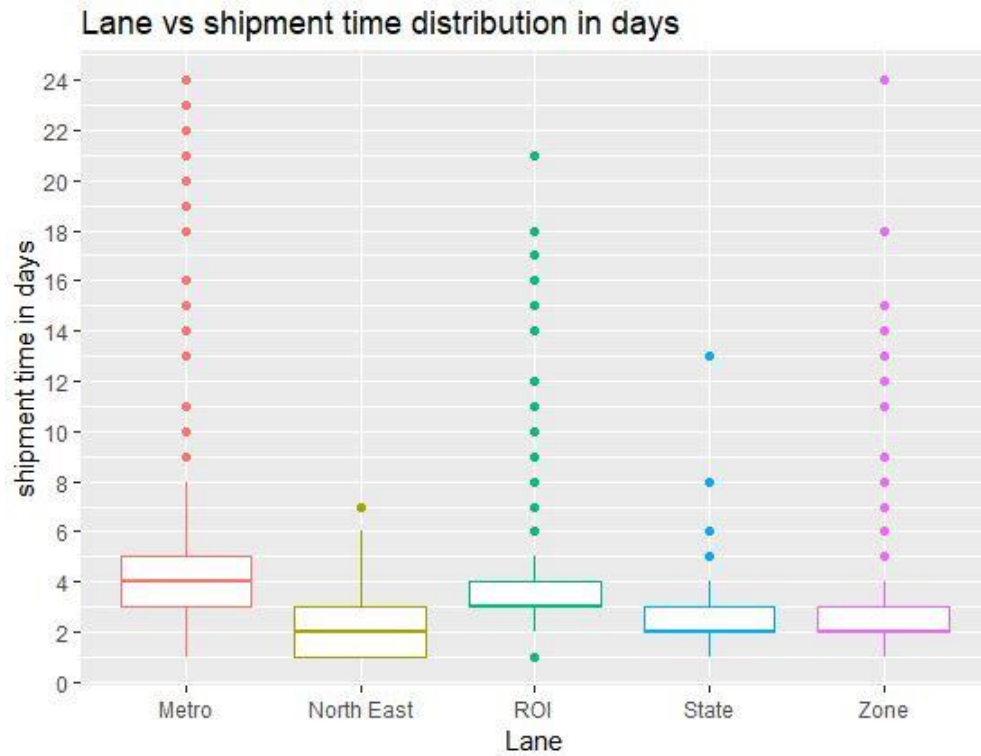


**The average processing time is 3.55 days**

Most of the shipments(97.9%) are delivered in 6 days.

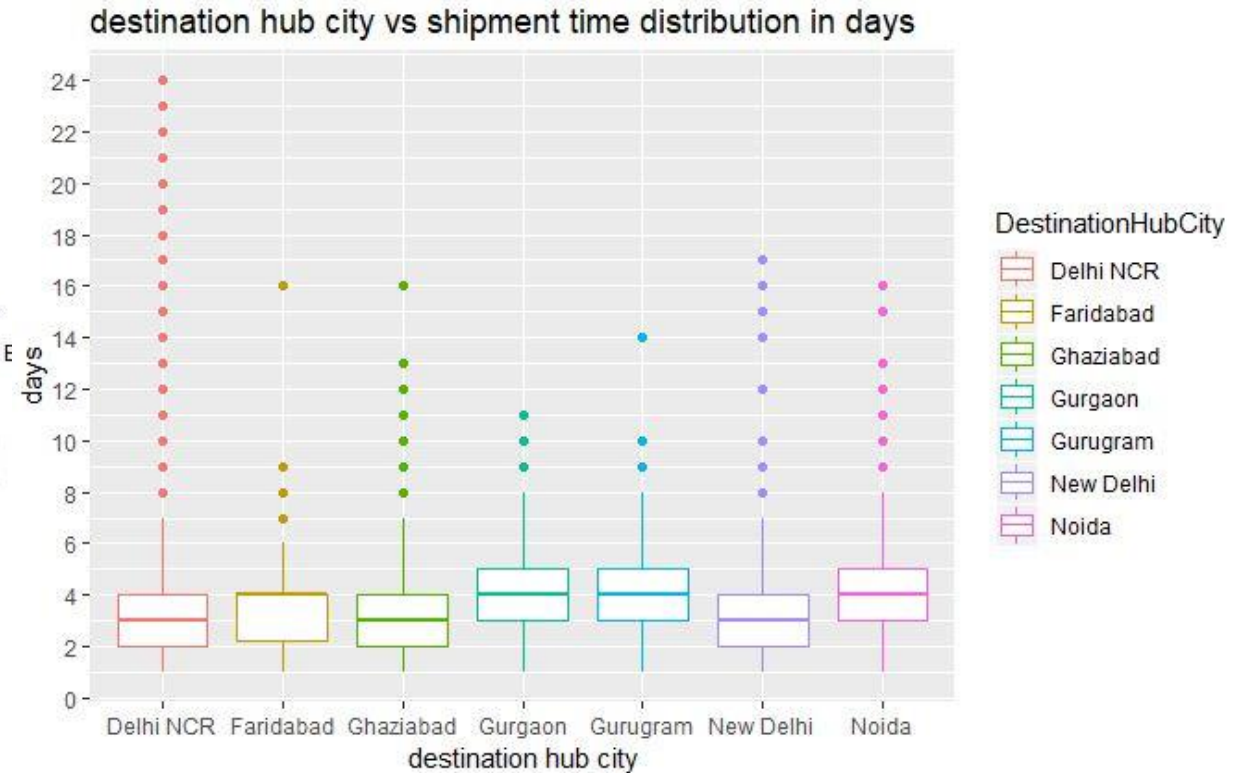
And only 2.1% shipments have taken more than 6 days(which are critical and need to analyse it).

# Shipping Time Distribution by Lane and Destination City



Avg. shipment time is maximum for Metro lane almost same for NE, State and Zone.

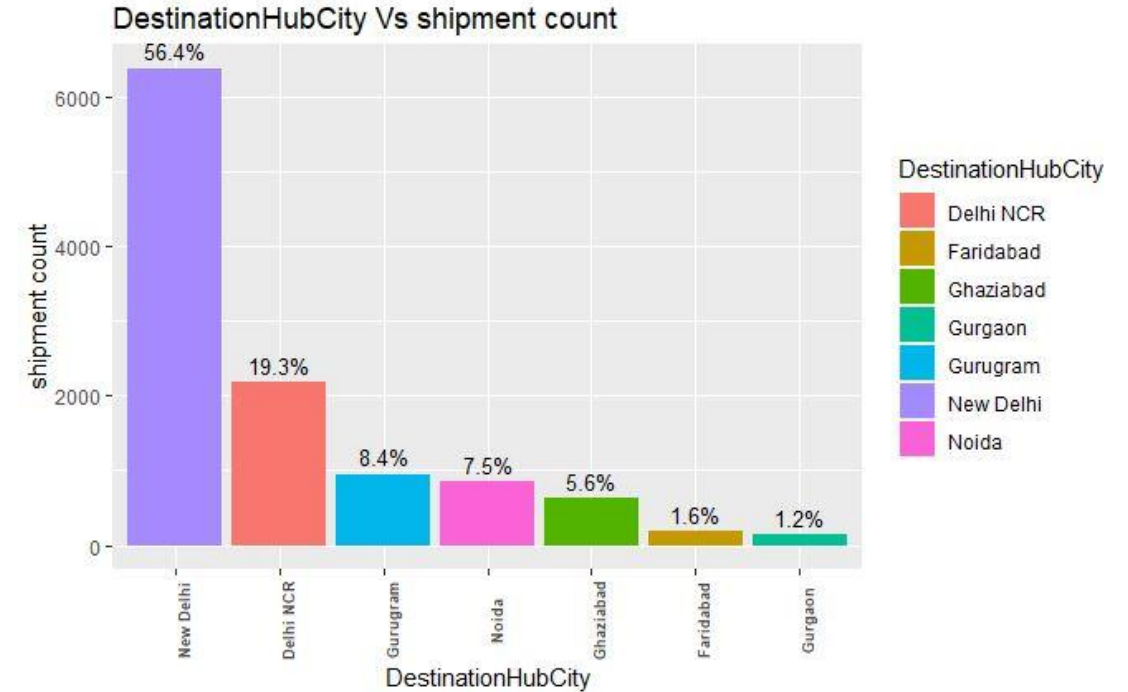
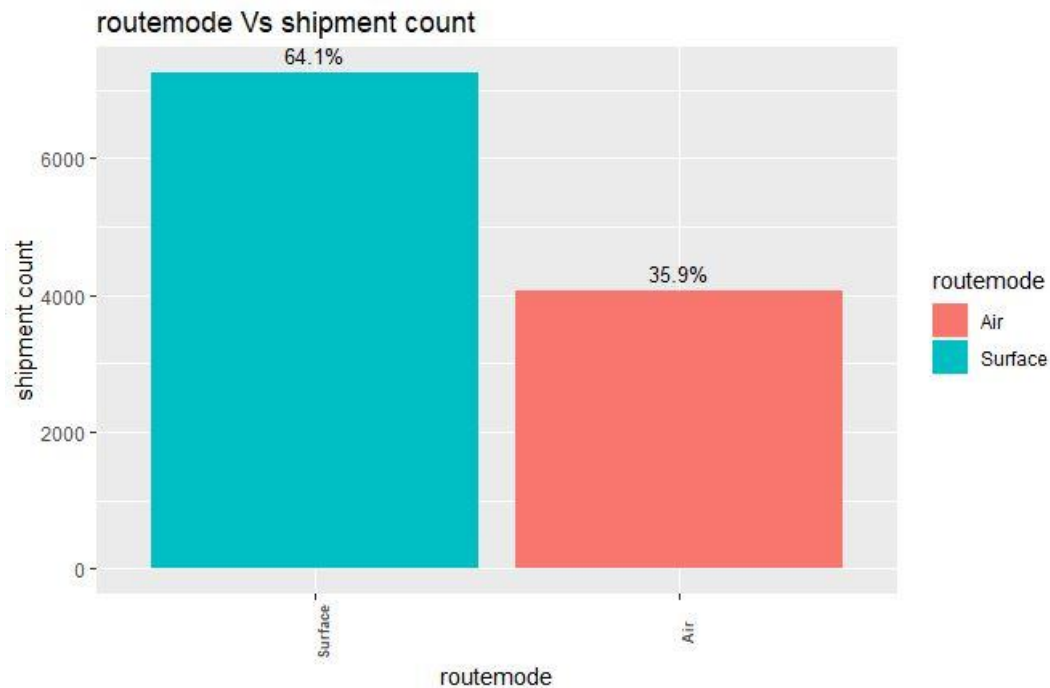
Overall avg. shipment time is less than 4 days (Is correct with reference to slide no 19).



Avg. shipment time highest for Gurgaon, Gurugram and Noida (all are neighbor locations).

Overall avg. shipment time is less than 4 days

## For highest shipment lane ROI (45.3%)

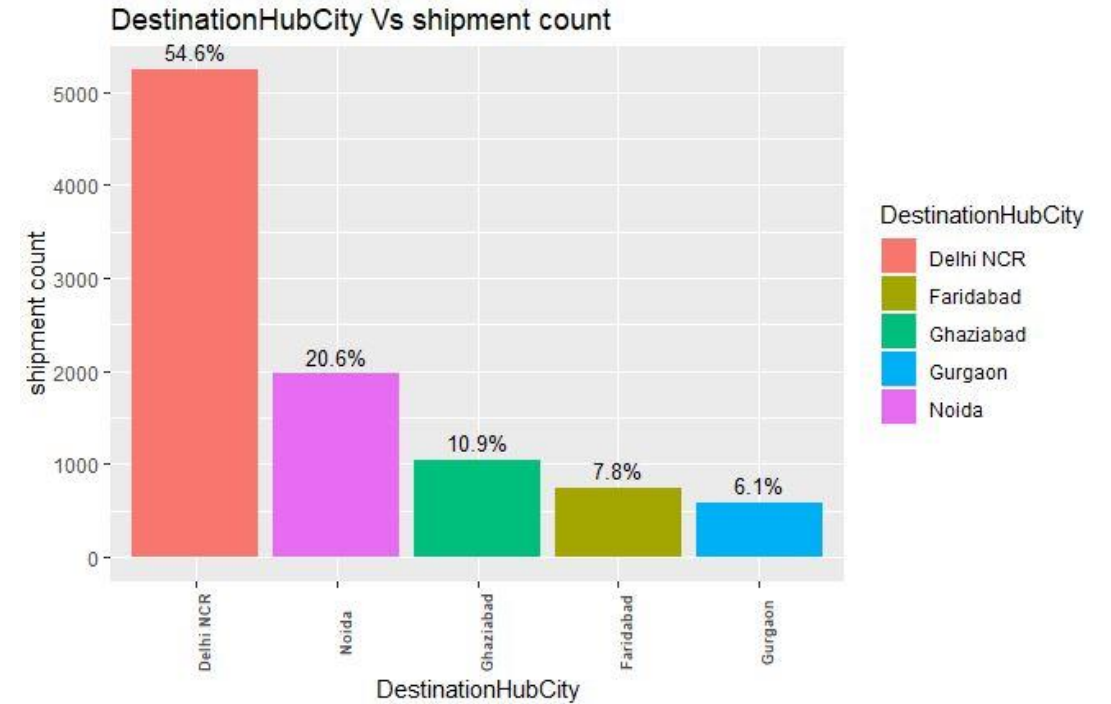
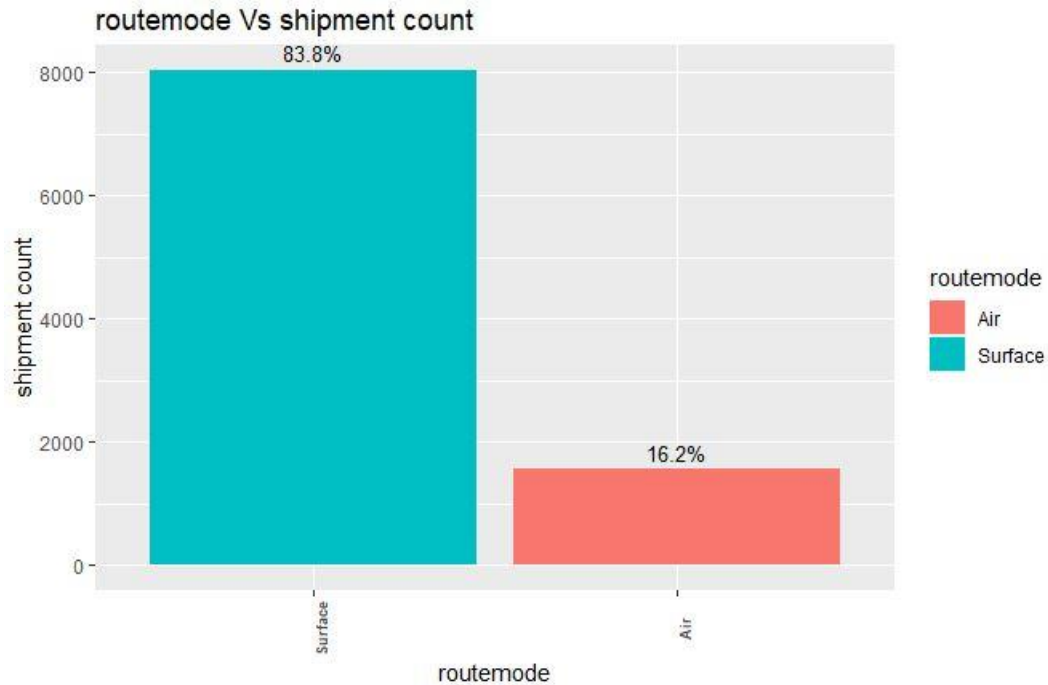


For ROI mean shipping time is 3.37 days.

Most number of shipments are happens from Bhiwandi – 31.8%, Surat – 19.2%, Bangalore – 18.7%, Pune – 16.1%.  
Most number of shipments destinations is New Delhi 56.4% , Delhi NCR – 19.3

Most shipment route is surface as destination shipping point could be away from airport, but there are good number of shipments happened by Air.

## For highest shipment lane Metro (38.4%)



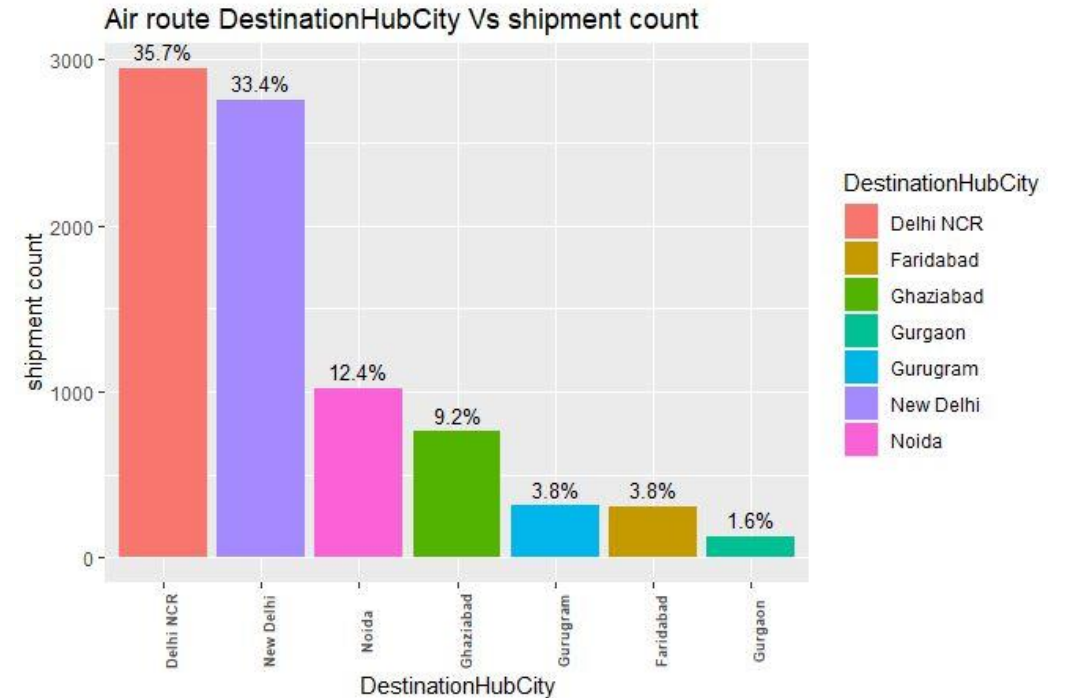
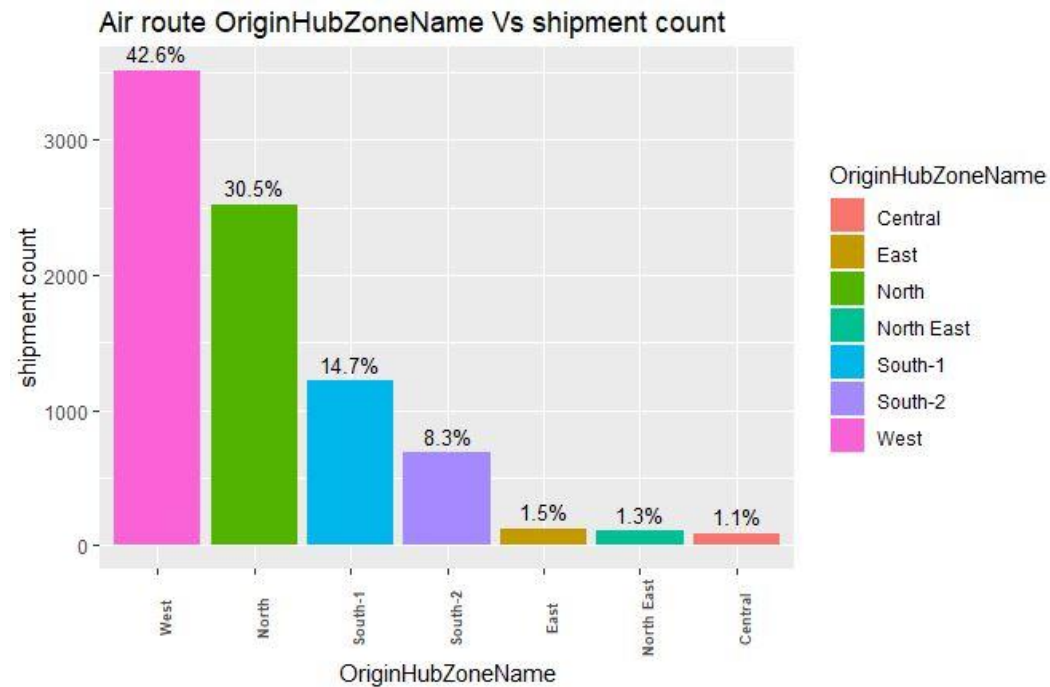
For Metro mean shipping time is 4.23 days

Most number of shipments are happens from Bangalore - 44.7%, Pune – 40.0%

Most number of shipments destination is Delhi NCR 54.4% , Delhi NCR – 20.6%

Most shipment route is surface because air transport is costliest and also most heavy shipments transportation happen by surface to reduce transportation cost which also takes more shipping time

# Route Mode - Air

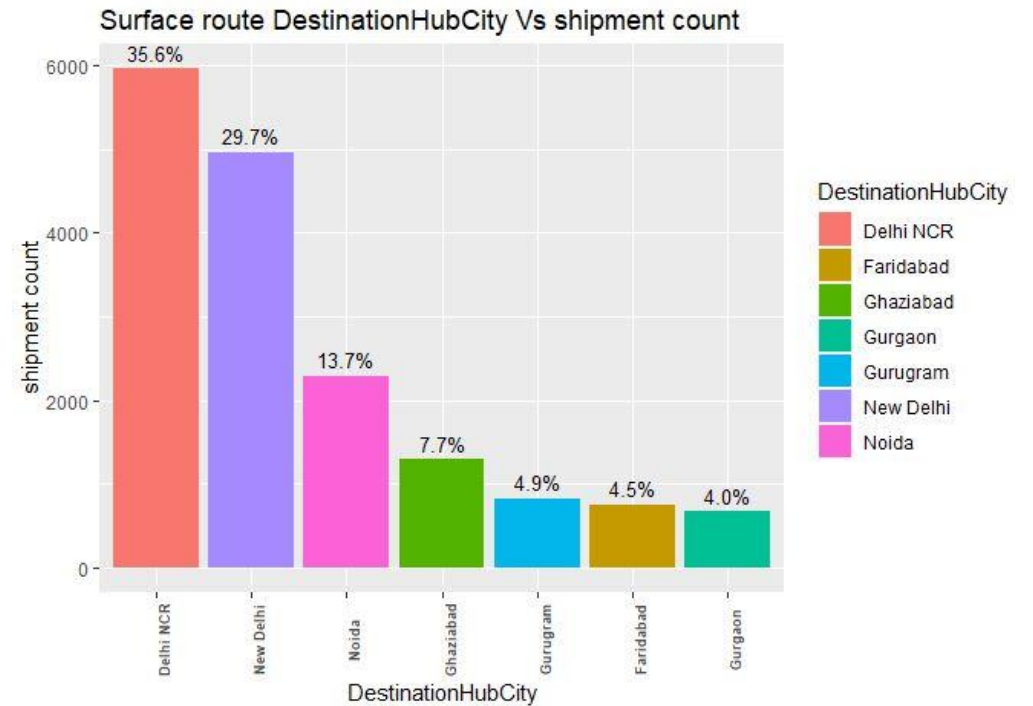
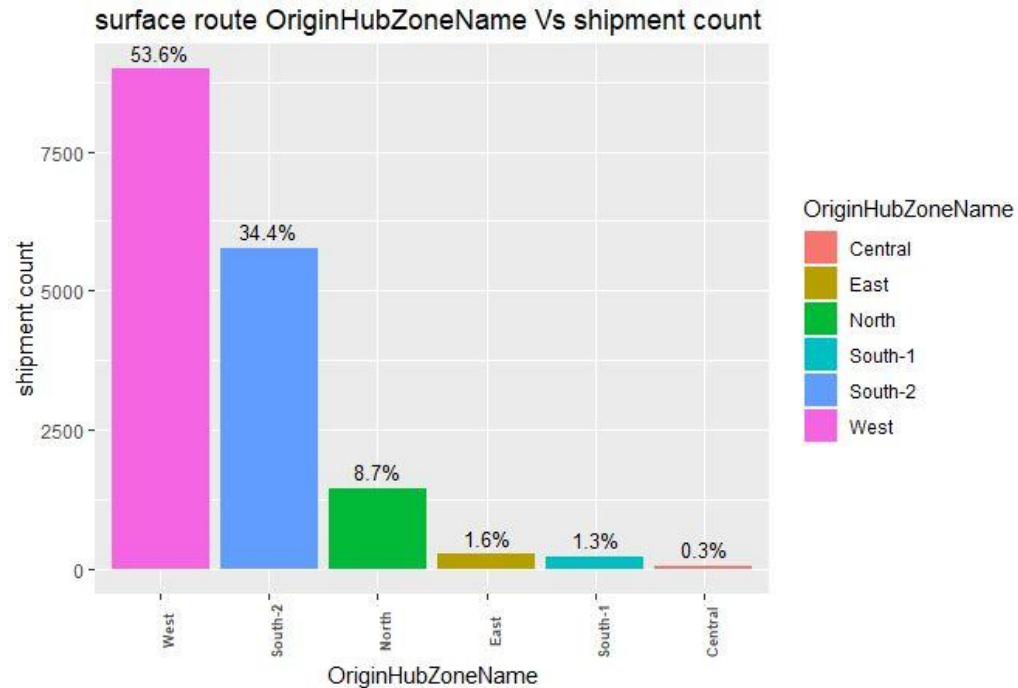


Avg. shipment time by air mode is 2.31 days

Highest shipment origin cities Bhiwandi – 22.4%, Chennai – 12%, Surat – 11.7%, Panipat – 9.4%, Bangalore – 8.1%, Ludhiana – 6.2%

Highest shipment destination cities are Delhi NCR, New Delhi and Noida

# Route Mode - Surface



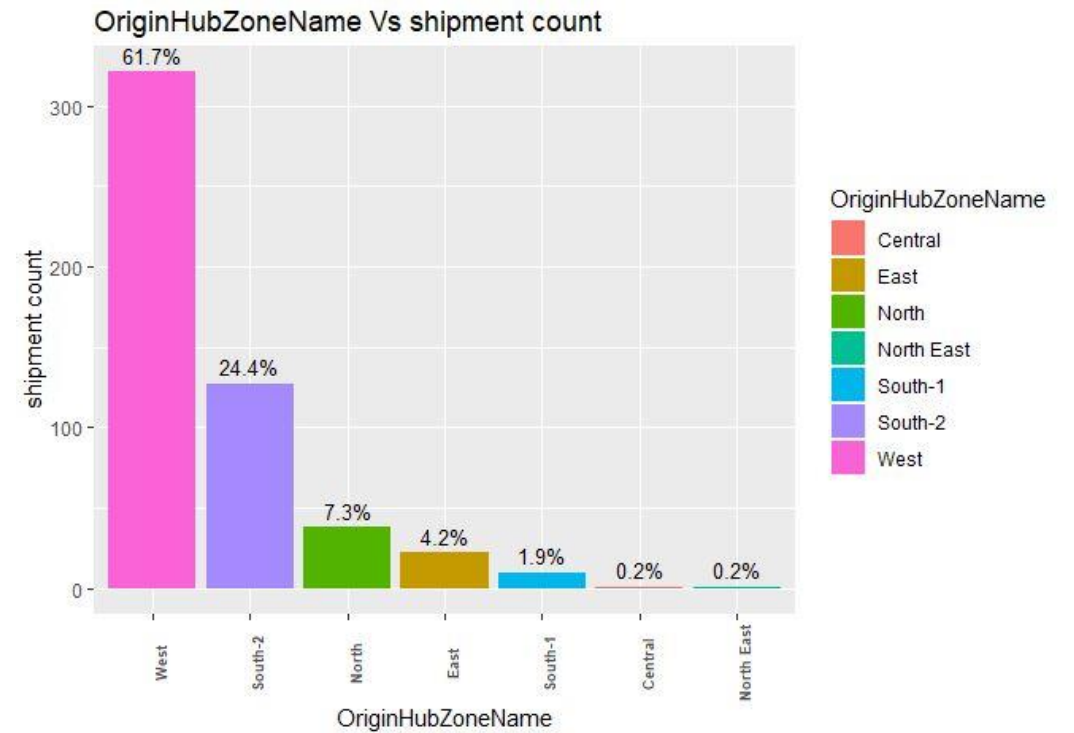
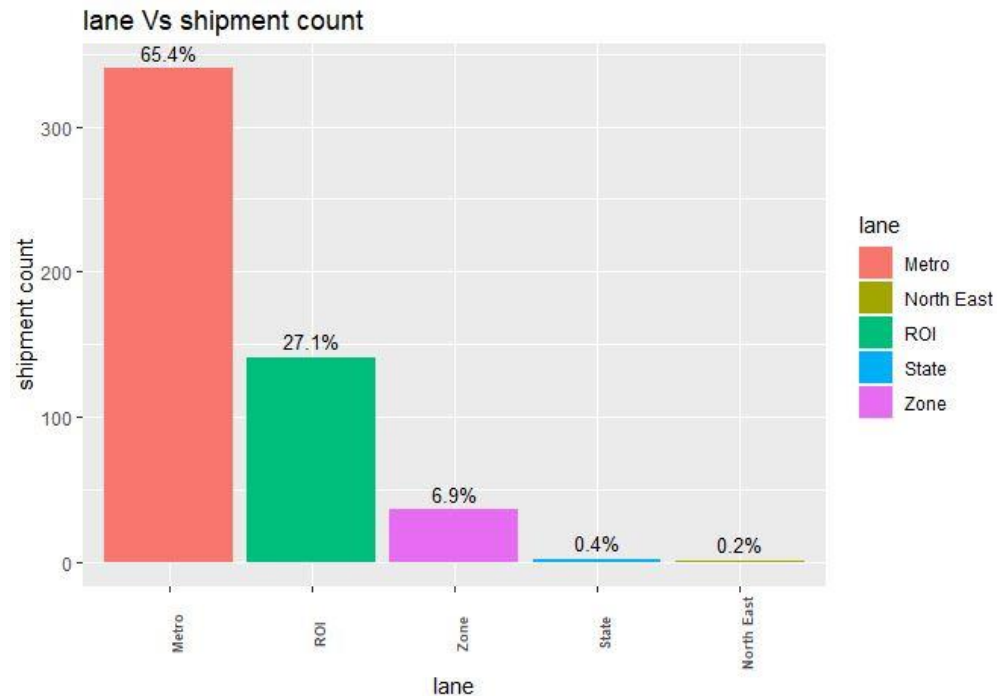
Avg. shipment time by surface mode is 4.16 days

Highest shipment origin cities Bangalore - 34.3%, Pune – 32.8% , Bhiwandi – 10.5% and Surat – 7.3%

Highest shipment destination cities are Delhi NCR, New Delhi and Noida & No shipment from North East



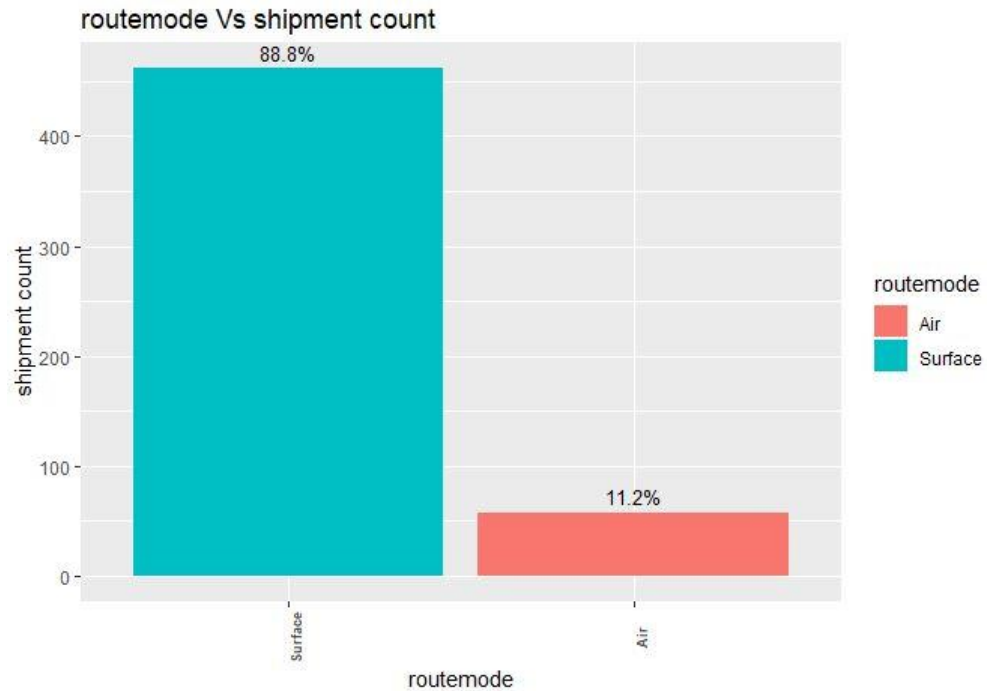
# Shipping Time more than 6 days



Reason to long time for shipping:

- Most of the shipment lane is between Metro and ROI, also 88.8% shipments route mode is surface(max travel time).
- Most of the shipments origin zone is West and South to North (Highest distance to be covered by surface)

# Shipping Time more than 6 days



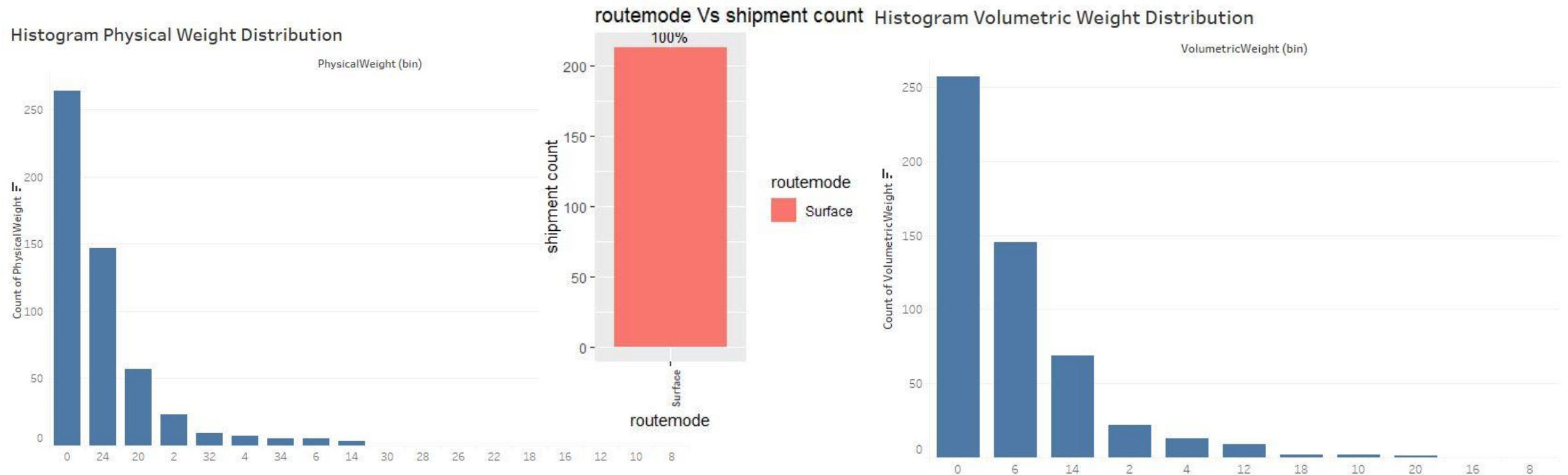
Average total shipment time(13.34d) here is 3.75 times more than overall shipping time(3.55d),

Also here time for processing at touch points and shipping between touch points has been increase by 1.5% than overall.

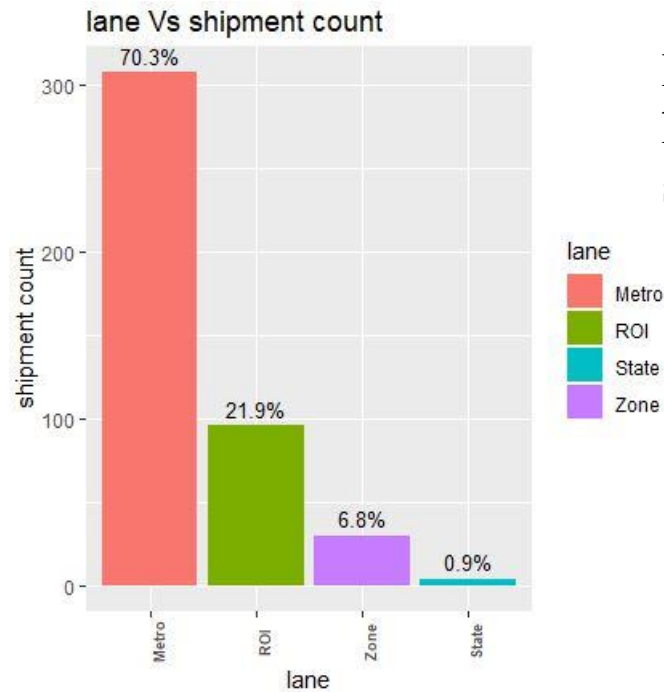
# Shipping Time more than 6 days between Processing Location 3 & 4

Reference to shipment between process location 3(Gurgaon) and 4(Delhi) for id's whose time difference is more 6 days are also part of the shipments with overall shipping time is more than 6 days

The reasons for this long shipment time is route mode and physical and volumetric weight of the shipments.  
between process location 3 and 4 here is 20 times more than overall shipping time (0.29 days)

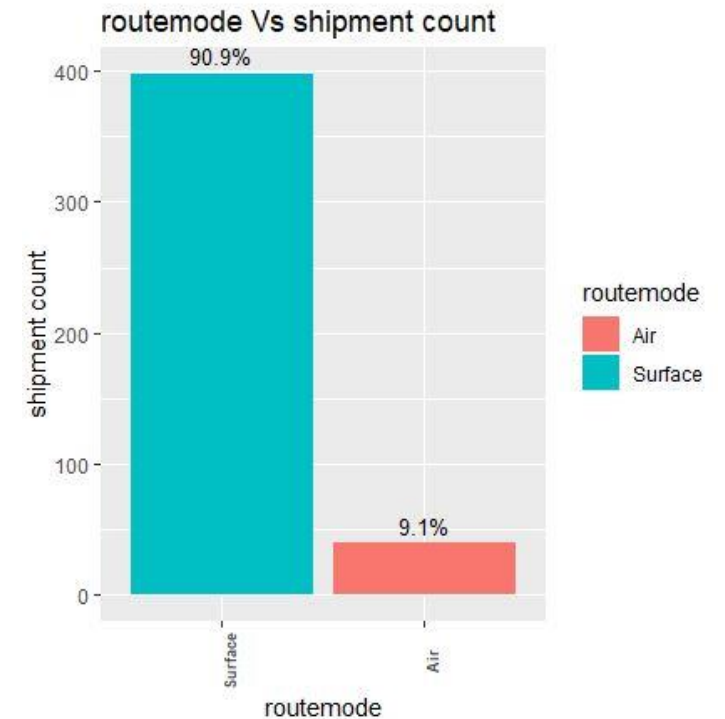


# Impact on Shipment of Physical Weight (above 4 kg)



For physical wt. above 4 kg shipments have taken highest time to delivered, more than 6 days to be shipped

North East Lane not available here

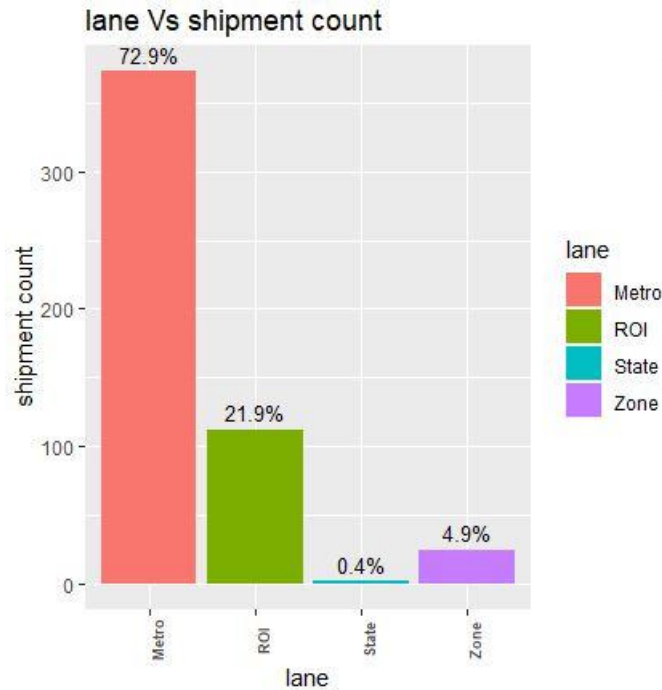


Most shipment lane is Metro and ROI by route mode surface due to distance in large, so more transit time.

And all heavy shipments transit mode is surface because shipping through air leads to more transportation cost.

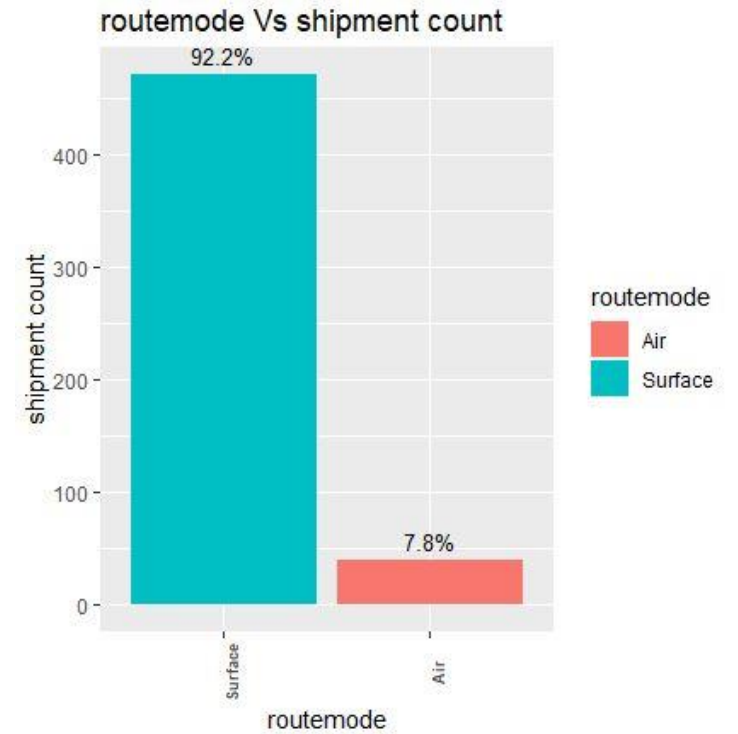
Also heavy shipments have been taking more time in processing at touch points/destination Points and leading to more day to shipped.

# Impact of Shipment on Volumetric Weight



For volumetric wt. above 5 kg shipments have taken highest time to delivered, more than 6 days to be shipped

North East Lane not available here



And all high volumetric weight shipment mode is surface, because shipping through air leads to more transportation cost.

**Service providers always choose weight which is higher between Physical or Volumetric and charges on it**

## Complicated Transit between hub/city route or travel mode & NULL value Analysis

- There are few shipments which has been process twice in same city OR in same state & then forwarded to destination OR few has taken Air mode for in zone transport
  - Cochin origin shipments has been process twice at Bangalore hub, route mode is Air
  - Mumbai origin shipments process twice at Bombay hub then again at Bhiwandi hub then sent to destination route mode is surface
  - Pune origin shipments 1<sup>st</sup> process at Pune then 2<sup>nd</sup> at Mumbai then sent to Delhi route mode is surface mode
  - Sonipat origin shipment 1<sup>st</sup> process at Sonipat, 2<sup>nd</sup> at Gurugram then 3<sup>rd</sup> again Sonipat and sent to Delhi
  - Vadodara origin shipment shipment 1<sup>st</sup> process at Vadodara, 2<sup>nd</sup> at Bhiwandi, 3<sup>rd</sup> at Mumbai then sent to Delhi route mode is Air
  - Yamuna Nagar origin shipment 1<sup>st</sup> process at Ambala, 2<sup>nd</sup> at Panipat then sent to Delhi when it is in same zone
- Some data observation 259 for NULL values data frame
  - Created new variable for total number of days for shipment
  - Converted default date format in standard format
  - All shipmen have been taken up to 6 days which meets our criteria (except one shipment).
  - Physical and Volumetric weight are below 6 kg/Cub.m
  - So not doing further analysis

# Observations from Analysis

- Most of the shipments (83.7%) are happening between **ROI** and **Metro lane**, its shipments with prefer route mode is surface and heavier shipments also transit by surface(61% of ROI + Metro).
- Average shipping time is same by destination city or Lane(3.55 days).
- For rout mode Air avg. shipping time is 2.31, and most of the shipments origin zones are West and North then most origin cities are Bhiwandi, Surat, Chennai, Panipat and Bangalore, ludhiana, Jaypur.
- For rout mode Surface shipping time is 4.16, and most origin zones are West and South then origin cities are Bangalore, Pune, Bhiwandi and Surat
- For both rout mode most of shipment destination cities are New Delhi, Delhi NCR and Noida.
- There are record/shipment Id's who has taken more than 6 days for shipment, and their reason are –
  - Most of the shipments lane is Metro and ROI, also 88.8% of them are transport by surface, so more travel time
  - Most of the shipments origin zone is West and South to North (Highest distance to be covered by surface)
  - Also average shipping time for those record is 5 times more than overall shipping time
  - And those records have 6 days (20% more) shipping time between process location 3(gurgaon) and 4(delhi) than overall shipping time 0.29 days between these points

# Observations from Analysis

- Shipments Physical and Volumetric weight have impact on the shipment time
  - It is observed that shipments have took more than 6 days to be shipped, when shipment weight is highest.
  - Here most of the shipment lane is Metro and ROI by route mode surface(distance between metros is in thousands km) so more transit time.
  - And all heavy shipments transit mode is surface because shipping through air leads to more
  - transportation cost. (transport through the surface is good option e.g.. railway or other vehicle transport ).
- Most of the shipments weight (65%) have physical less than or equal to 0.5 kg, and shipments with weight (57%) volumetric less than equal to 0.5 cubic meter.
- Shipment charges are more when considered the volumetric weight and for Air transit.
- Some shipments transit route or processing locations can be optimize to reduce time and cost
  - Like same zone shipments can be sent by surface and with least processing than processing at each/same processing time
  - Few shipments are processed thrice in same state then sent to destination location, here we can reduce time and travelling cost also. Eg. Shipment origin Mumbai and Delhi process most of the time in between same cities/state From Bhiwandi which are supposed to delivered by Air, those only need to process at Mumbai



# Key Performance Indicator

- Time to process shipment at processing locations
- Time for shipment between processing location/Touch points
- Actual shipment time (for each shipment, By lane and destination hub)
- Cost of transport (Based on weight and route mode)
- Route of shipments (Air or Surface)
- Physical and Volumetric weight

# Suggestion And Recommendation

- Most of the destination are near Delhi then we can keep common final processing location at centre location
- Shipments which have long distance travel then for them processing/touch point can be reduce for both Air and Surface mode
- 75% Shipment origin zone are West and South (Pune, Bangalore, Bhiwandi / Mumbai) to the North(New Delhi, Delhi NCR and Noida). In which for most route mode is Surface, so we can reduce processing location from 4 to 3 or 2
- Processing location 1 has highest processing time as one reason could be that all heavy weight shipments takes more processing time and we need to be investigate other reason as we don't know what happening there.
- 60% shipments 2<sup>nd</sup> processing location is DEL/Gurgaon and then also again 3<sup>rd</sup> and 4<sup>th</sup> processing location is Delhi, so here we can directly transport to 4<sup>th</sup> location which will reduce cost and time also. And this is the reason for shipment time between processing location 1 and 2 is highest
- Surface average shipping time is 4.16 days, which can be reduced if we reduced possible process location (with reference to above two points)
- We can reduce processing time by processing same order in same city or hub more than once and only shipments who are supposed to transit by air can be processed at Metro location where Airport facility is near (Like Bhiwandi shipments can be processed at Mumbai as route mode is Air)
- If reducing processing location saving enough cost and if possible then to reduce time we can choose air mode
- Here we don't know what kind of goods is transporting, but perishable product can be sent through Air as average air shipping time is less than surface and other goods can be sent through surface if time is not concern

## Suggestion And Recommendation

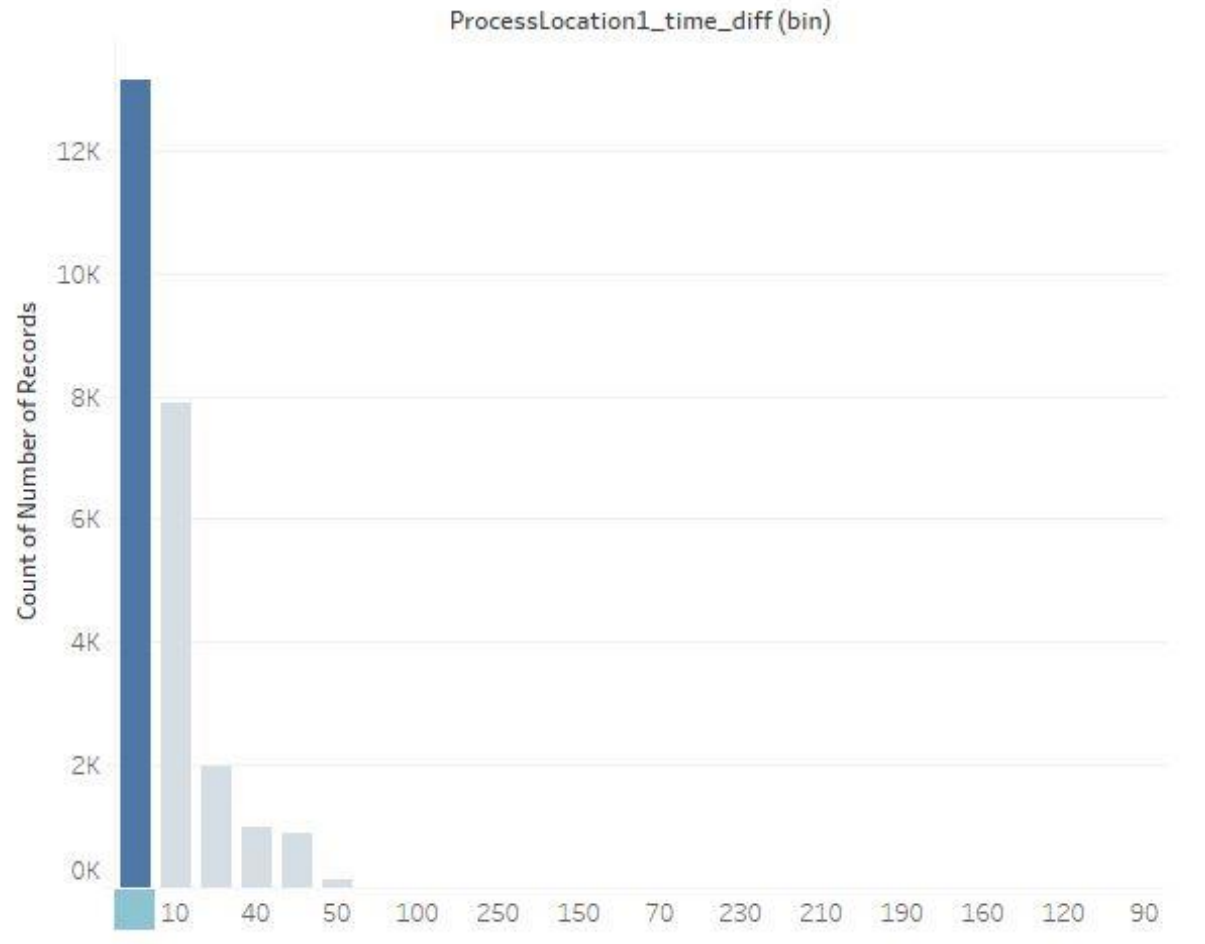
- For shipments have been taken more than 6 days, heavy physical /volumetric weight and route mode surface are reason, but it is also possible that heavy shipments shipping time is more because of the conditions like long distance and to prevent damage to goods
- Only heavy/ high volume shipments can be transit by surface to reduce high air shipping fare, but priority shipments can be sent by Air

Thank You

Few more points to get idea about what  
happened between touch points

# Shipment Processing Time at Processing Location/Touch Point 1

Process Location 1 Time Processing Distribution in Hour

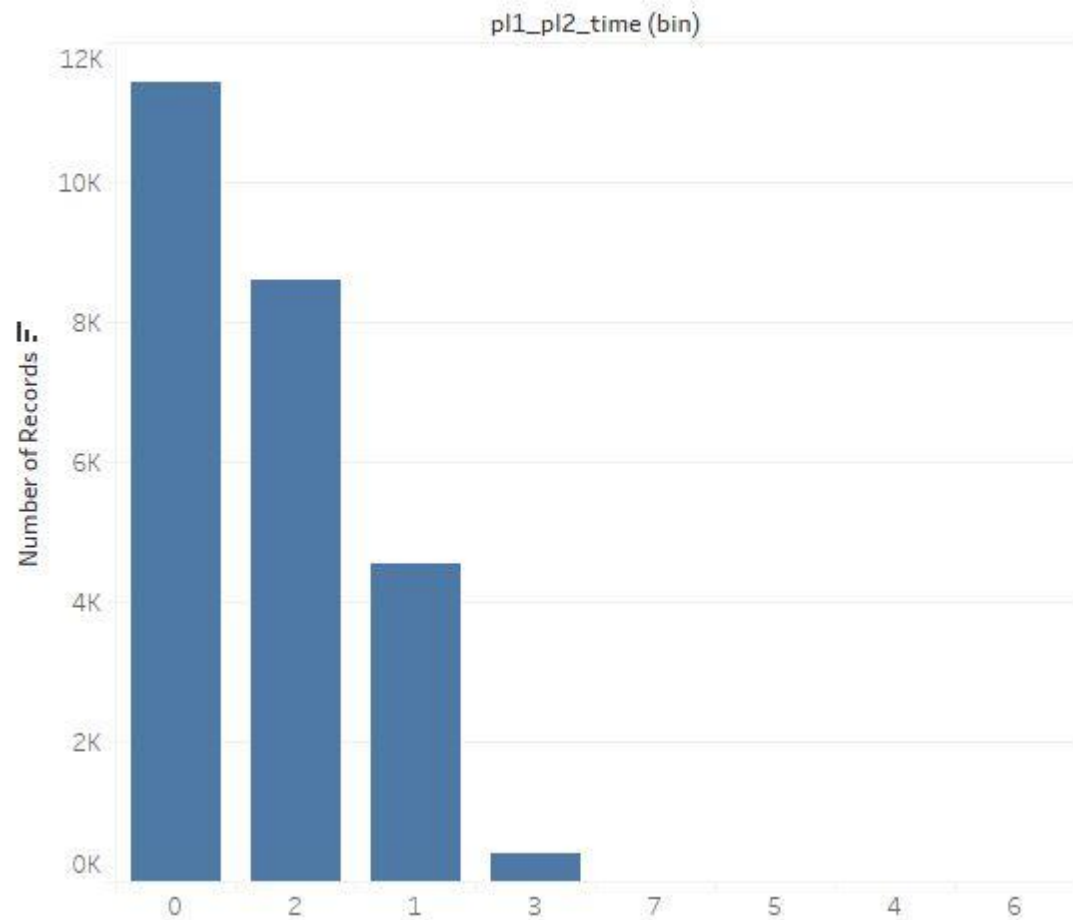


Here for most of the shipments (98.72%) processing time is between 0 to 48 hrs. And very few shipments(1.27%) have taken more than 48 hours.

**The average processing time is 12.53 hours**

# Shipment Time between Processing Location/Touch Point 1 and 2

Shipment Time Distribution Between Touch Point 1 and 2 in Days

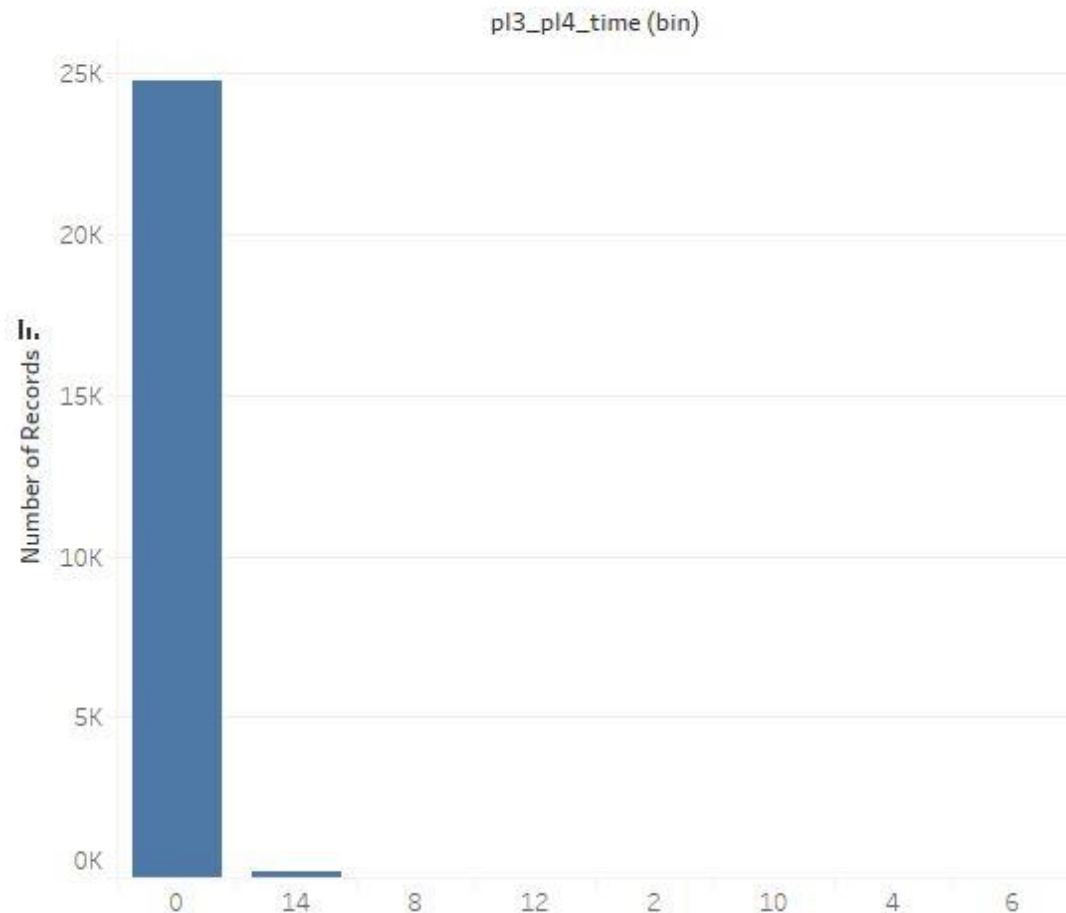


Here for most of the shipments (98.31%) shipped within 3 days.  
And very few shipments(1.68%) have been shipped in more than 3 days.

**The average shipping time is 1.29 days**

# Shipment Time between Processing Location/Touch Point 3 and 4

Shipment Time Distribution Between Touch Point 3 and 4 in Days



Here for most of the shipments (99.05%) shipped within 1 day  
And very few shipments(0.94%) have been shipped in more than 1 days.

**The average shipping time is 0.29 day (less than half day)**

**Here last bin represents no. of shipments who have taken 14 days to transit which have explained further**