



# Machine Learning Project

Transportation Rate Prediction Model

Teams: Peter Tian, Shuyang Lin, Weixi Chen  
October, 2022

Support By:





# Machine Learning Project

Transportation Rate Prediction Model

## Background

Group	Deutsche Post DHL Group			
	One global team			
	Lean corporate center			
Corporate Divisions	Post – eCommerce – Parcel	Express	Global Forwarding Freight	Supply Chain
Brands <sup>1</sup>	Deutsche Post			
	DHL Customer Solutions & Innovation			
	Selective group wide shared functions and services			

### DHL SUPPLY CHAIN SOLUTIONS



#### Transport Solutions

DHL Managed Transport & Brokerage

DHL Inbound Transport Operations

DHL Outbound Transport Operations



#### Warehousing Solutions

DHL Warehousing

DHL Real Estate Solutions



#### Management Services

DHL Logistics Consulting

DHL Inventory Financing Services

DHL Procurement Services

DHL Recall Services

DHL Business Support Center



#### Integrated Solutions

DHL Lead Logistics Partner

DHL Inbound to Manufacturing

DHL Service Logistics

DHL Indirect Materials Management (MRO)

DHL Passenger Gateway Services

DHL Envirosolutions

DHL HealthConnect

DHL Clinical Trials



#### Additional Services

DHL Packaging Services

DHL Pre-Sales Services

DHL In-plant Logistics



## Market leader in Contract logistics

# #1

Contract Logistic Provider

Managing Supply Chains to reduce complexities



50+

Countries Covered



~160 M

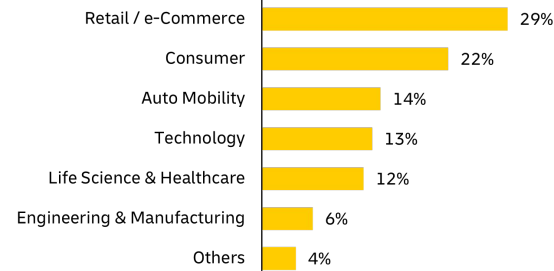
Sq. feet of Storage Space



~EUR 14 bn

Revenue

## 2021 Revenue Mix



Supported By:





## Project Overview

DHL looks third-party logistic trucks for :



TruckLoad Shipment (TL)

DHL looks third-party logistic trucks for:



Less-Than-TruckLoad Shipment (LTL)

### DHL Transport Solution Team:

- The team quotes third-party companies contract prices or spot prices to find feasible transportation solutions for client's shipments
- The team considers quoted price in different data frames as a benchmark to find lowest price
- The team will save quoted price and historical shipments in the database

### With available historical shipment data and quoted price data:

A model is seeked to **predict future transportation rates** by inputting available shipment information and quoted price.



## Dataset

Data is provided by DHL Operation and Analytics Team. All data is for study use and has been masked

### Historical Shipment Data (MT\_Data)

**Description:** The dataset includes 4 month historical shipment information

**Column:** 31

**Rows:** 4,963,508

**Column Names:**

```
[Unnamed: 0.1], 'SHIPMENT ID', 'CUSTOMER', 'DISTANCE',  
'CASES', 'WEIGHT', 'VOLUME', 'SOURCE LOCATION ID', 'ORIGIN NAME',  
'ORIGIN CITY', 'ORIGIN STATE', 'ORIGIN ZIP', 'DEST LOCATION ID',  
'CONSIGNEE NAME', 'DEST CITY', 'DEST STATE', 'DEST ZIP',  
'ACTUAL CARRIER', 'ACTUAL MODE', 'ACTUAL EQUIP', 'LINEHAUL COSTS',  
'FUEL COSTS', 'ACC. COSTS', 'TOTAL ACTUAL COST', 'PU_APPT', 'DL_APPT',  
'PU_ARRIVAL (X3)', 'PU_DEPARTED (AF)', 'DL_ARRIVAL (X1)',  
'DL_DEPARTED (D1)', 'Insert Date']
```

### Historical Quoted Price Data (DAT\_Data)

**Description:** The dataset includes 4 month historical quoted price data information

**Column:** 40

**Rows:** 63,486

**Column Names:**

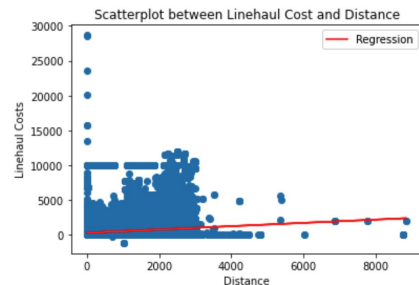
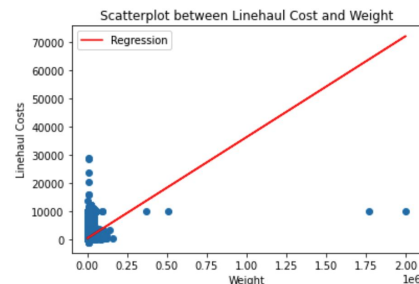
```
['PC_MILER_PRACTICAL_MILEAGE', 'SPOT_AVG_LINEHAUL_RATE',  
'SPOT_LOW_LINEHAUL_RATE', 'SPOT_HIGH_LINEHAUL_RATE',  
'SPOT_FUEL_SURCHARGE', 'SPOT_TIME_FRAME', 'SPOT_ORIGIN_GEO_EXPANSION',  
'SPOT_DESTINATION_GEO_EXPANSION', 'SPOT_NUMBER_OF_COMPANIES',  
'SPOT_NUMBER_OF_REPORTS', 'SPOT_LINEHAUL_RATE_STDDEV',  
'SPOT_YOUR_OWN_AVG_LINEHAUL_RATE', 'SPOT_YOUR_OWN_NUMBER_OF_REPORTS',  
'SPOT_ERROR', 'CONTRACT_AVG_LINEHAUL_RATE',  
'CONTRACT_LOW_LINEHAUL_RATE', 'CONTRACT_HIGH_LINEHAUL_RATE',  
'CONTRACT_FUEL_SURCHARGE', 'CONTRACT_AVG_ACCESSORIAL_EXCLUDES_FUEL',  
'CONTRACT_TIME_FRAME', 'CONTRACT_ORIGIN_GEO_EXPANSION',  
'CONTRACT_DESTINATION_GEO_EXPANSION', 'CONTRACT_NUMBER_OF_COMPANIES',  
'CONTRACT_NUMBER_OF_REPORTS', 'CONTRACT_LINEHAUL_RATE_STDDEV',  
'CONTRACT_YOUR_OWN_AVG_LINEHAUL_RATE',  
'CONTRACT_YOUR_OWN_NUMBER_OF_REPORTS', 'CONTRACT_ERROR', 'ORIG_CITY',  
'ORIG_STATE', 'ORIG_POSTAL_CODE', 'DEST_CITY', 'DEST_STATE',  
'DEST_POSTAL_CODE', 'TRUCK_TYPE', 'ID_OR_OPTIONAL_FIELD', 'NOTE',  
'RUN_ID', 'SOURCE_FILE_NAME', 'DWH_INSERT_DATE']
```



VANDERBILT



## Correlation Map:



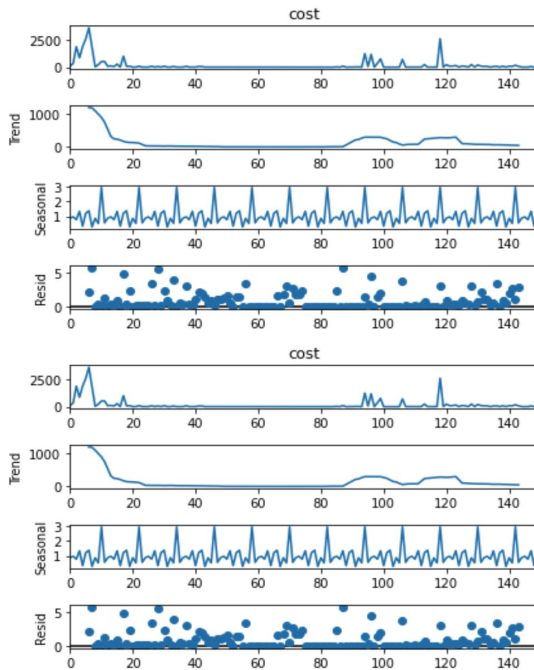
**Supported By:**





## Basic EDA Exhibit

### Sample Linehaul Cost Time Series Analysis:



Based on Sample MT Data

Augmented Dickey-Fuller

Fuller Results

Test Statistic -4.713

P-value 0.000

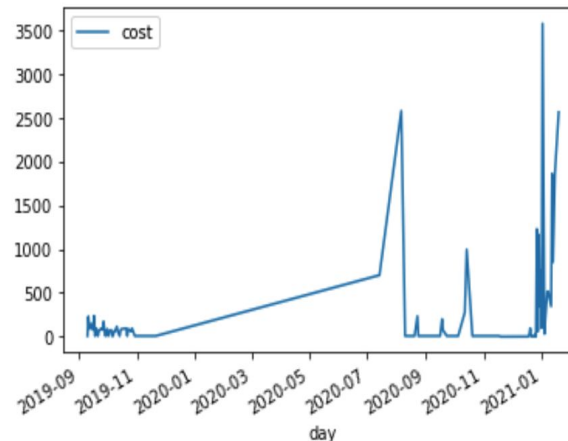
Lags 1

Trend: Constant

Critical Values: -3.48 (1%), -2.88 (5%), -2.58 (10%)

Null Hypothesis: The process contains a unit root.

Alternative Hypothesis: The process is weakly stationary.



Supported By:

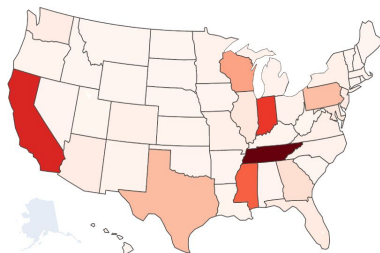




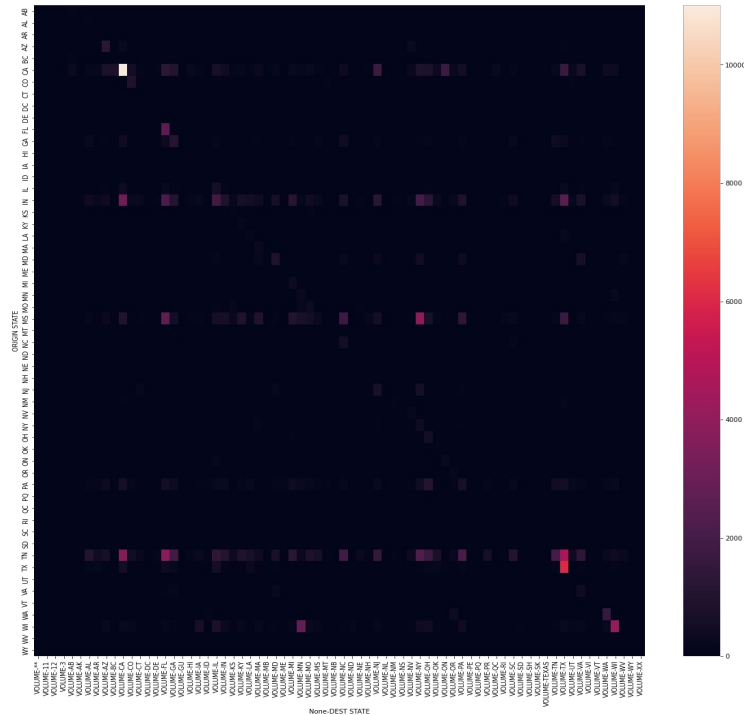
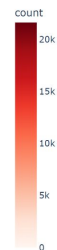
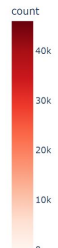
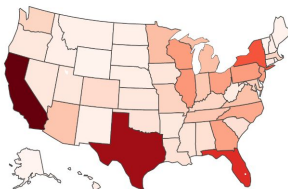
## Basic EDA Exhibit

### Shipment Map:

cases count of different origin state



cases count of different destination state



Based on Sample MT Data

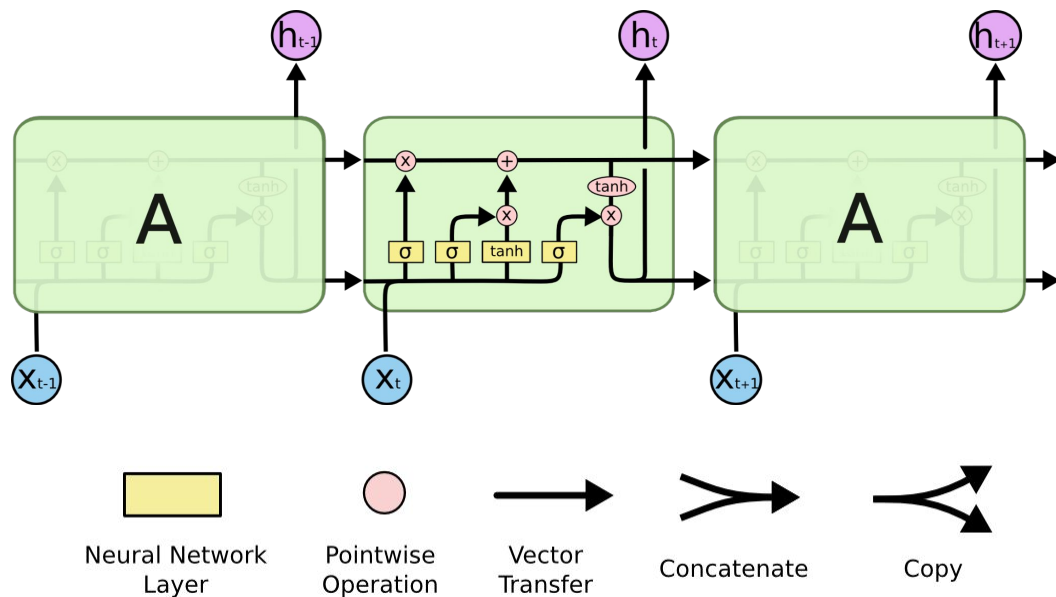
Supported By:





## Potential Model Selection

Long Short-Term Memory - a special kind of Recurrent Neural Network



Good to carry both long term and short term info by:

- Cell state vector carrying memories
- Gates deciding to forget or not
- Tanh layers to process candidates