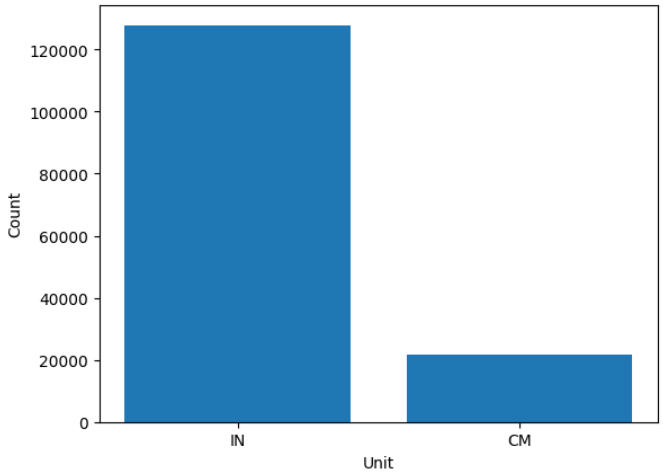
Table of Content

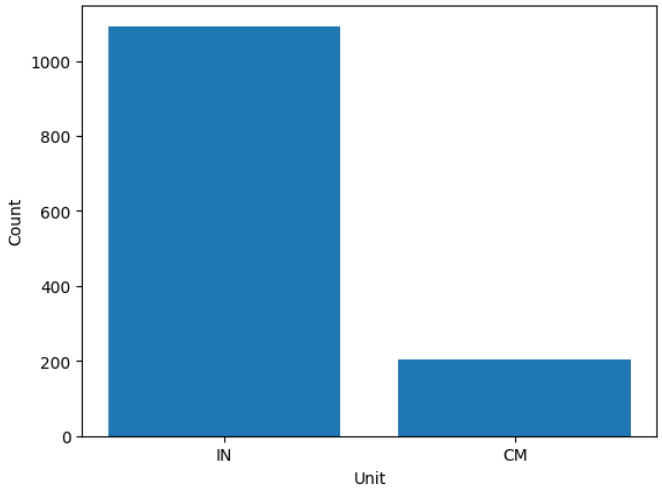
Links

- Colab notebook(Code/Script):
  Incorrect Package Dimensions Detection Model.ipynb
- Google Sheet:
  Incorrect Package Dimensions Detection
- Data:
  Incorrect Package Dimensions Detection
- BlueTick AI Micro Experience:
  Incorrect Package Dimensions Detection Model

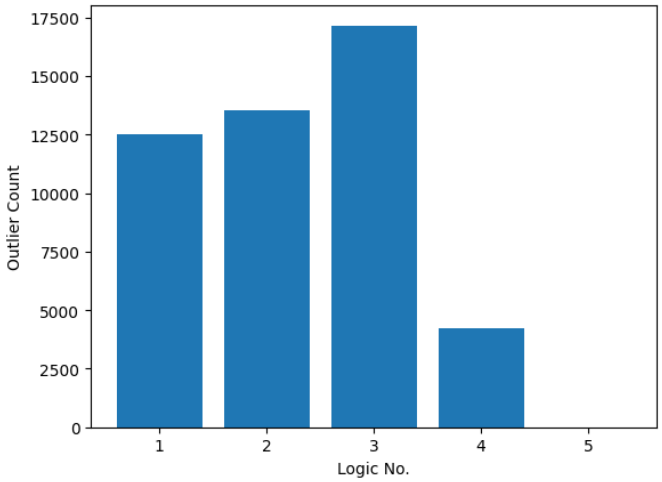
Plots



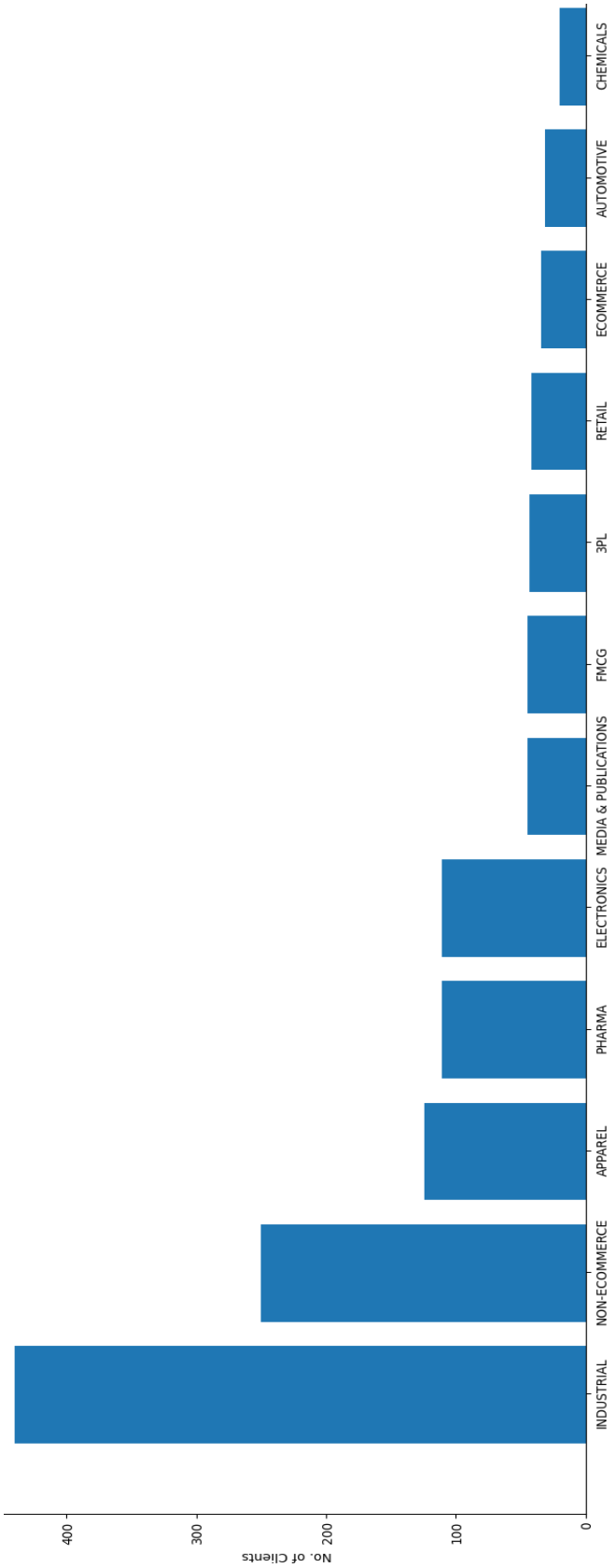
1. Count of inches against centimeters in consignment_volume_data



2. Number inch clients against number of cm clients for output_2



3. Outliers count using different logics



4. Number of Clients for each industry

Views of DataFrames

	id	created_date	cnote	client_id	weight	Voloume	total_boxes	industry_type	QC_Validation	delivered	Delivery_Date	cpm
0	32123	15-01-2019 00:00	8000522895	1112	13.0	3.23	1	APPAREL	Yes	1	19-01-2019 17:01	1
1	32124	15-01-2019 00:00	8000119761	1113	122.0	23.30	14	3PL	Yes	1	17-01-2019 11:18	1
2	32125	15-01-2019 00:01	8000522921	1112	19.0	3.23	1	APPAREL	Yes	1	18-01-2019 10:38	1
3	32126	15-01-2019 00:01	8000574330	1114	68.0	11.16	4	3PL	Yes	1	18-01-2019 17:42	0
4	32127	15-01-2019 00:02	1002478259	1115	1075.0	73.35	50	CHEMICALS	Yes	1	22-01-2019 17:29	1
5	32128	15-01-2019 00:02	8000522929	1112	50.0	6.46	2	APPAREL	Yes	1	16-01-2019 12:58	1
6	32129	15-01-2019 00:02	1002229812	1116	354.0	42.50	34	MEDIA & PUBLICATIONS	Yes	1	24-01-2019 21:17	1
7	32130	15-01-2019 00:03	1002264337	1117	380.0	285.97	95	APPAREL	Yes	1	18-01-2019 17:01	1
8	32131	15-01-2019 00:03	8000522922	1112	19.0	3.23	1	APPAREL	Yes	1	19-01-2019 14:25	1
9	32132	15-01-2019 00:04	1002357239	1118	50.0	21.44	10	INDUSTRIAL	Yes	1	18-01-2019 17:50	1

1. View of consignment_data

	consignment_id	created_date	cnote	client_id	weight	volume	total_boxes	industry_type	QC_Validation	delivered	Delivery_Date	cpm
0	32123	15-01-2019 00:00	8000522895	1112	13.0	3.23	1	APPAREL	Yes	1	19-01-2019 17:01	1
1	32124	15-01-2019 00:00	8000119761	1113	122.0	23.30	14	3PL	Yes	1	17-01-2019 11:18	1
2	32125	15-01-2019 00:01	8000522921	1112	19.0	3.23	1	APPAREL	Yes	1	18-01-2019 10:38	1
3	32126	15-01-2019 00:01	8000574330	1114	68.0	11.16	4	3PL	Yes	1	18-01-2019 17:42	0
4	32127	15-01-2019 00:02	1002478259	1115	1075.0	73.35	50	CHEMICALS	Yes	1	22-01-2019 17:29	1

2. View of consignment_data after renaming columns to correct value

	consignment_id	client_id	weight	volume	total_boxes	industry_type
0	32123	1112	13.0	3.23	1	APPAREL
1	32124	1113	122.0	23.30	14	3PL
2	32125	1112	19.0	3.23	1	APPAREL
3	32126	1114	68.0	11.16	4	3PL
4	32127	1115	1075.0	73.35	50	CHEMICALS

3. Visualization of P2_consignment_data after dropping irrelevant columns

	consignment_id	length	breadth	height	unit	number_of_boxes	created_at
0	32123	31.0	12.0	15.0	IN	1	NaN
1	32124	22.0	8.0	12.0	IN	2	NaN
2	32124	23.0	13.0	11.0	IN	4	NaN
3	32124	20.0	11.0	13.0	IN	8	NaN
4	32125	31.0	12.0	15.0	IN	1	NaN

4. Visualization of P2_consignment_volume_data

	consignment_id	length	breadth	height	unit	number_of_boxes
0	32123	31.0	12.0	15.0	IN	1
1	32124	22.0	8.0	12.0	IN	2
2	32124	23.0	13.0	11.0	IN	4
3	32124	20.0	11.0	13.0	IN	8
4	32125	31.0	12.0	15.0	IN	1

5. View of consignment_volume_data after dropping useless column

Tables

1. Sample data of iqr_analysis of consignment_data

Client ID	Count of Consignments	Total Weight	Total Boxes	Mean CFT	25 %ile CFT	75 %ile CFT	IQR-Lower Limit	IQR-Upper Limit	Total Volume
1112	1186	36762.22	2145	5.95183663	4.811661507	6.857142857	1.743439481	9.925364883	6555.87
1113	51	3804.6	201	8.136285479	5.149336049	8.613375864	-0.046723674	13.80943559	588.76
1114	503	32251	2064	7.352325403	5.902608413	7.990599295	2.77062209	11.12258562	5070.25
1115	91	52530.8	2797	55.33908825	10.44487124	17.58665821	-0.267809222	28.29933868	3727.19
1116	781	129519.87	9791	10.337240 3	7.466666667	12.1	0.516666667	19.05	13675.42
1117	68	32093.76	2736	4.102478128	1.725664743	5.366554863	-3.735670439	10.82789004	8229.69
1118	9	594	80	2.920151418	2.587800337	3.222094361	1.636359382	4.173535349	413.03
1119	12	6131.32	218	4.738853504	3.862019808	4.064526656	3.558259534	4.36828693	1520.67
1120	36	8298.5	590	5.837575005	5.2249054	6.246474423	3.692551865	7.778827958	1451.76

Description:

Client ID	Ids of clients
Count of Consignments	Total no. of consignments for a client
Total Weight	Total weight of consignment for the client
Total Boxes	Total no. of boxes for the client
Mean CFT	Mean of densities consignments for the client
25 %ile CFT	25% quantile of densities for the client
75 %ile CFT	75% quantile of densities for the client
IQR-Lower Limit	IQR lower limit of densities for the client
IQR-Upper Limit	IQR upper limit of densities for the client
Total Volume	Total volume of consignments for the client

2. Sample data of processed consignment_volume_data

consignment_id	length	breadth	height	unit	number_of_boxes
32123	31	15	12	IN	1
32124	22	12	8	IN	2
32124	23	13	11	IN	4
32124	20	13	11	IN	8
32125	31	15	12	IN	1
32126	132	122	12	CM	1
32126	130	27	20	CM	1
32126	38	34	34	CM	1
32126	33	22	12	CM	1

Description:

consignment_id	Every consignment id
length	Length for each consignment
breadth	Breadth for each consignment
height	Height for each consignment
unit	Unit for each consignment either in inch or cm
number_of_boxes	Number for each consignment

3. Sample data of analysis of consignment_volume_data

Client ID	Industry Type	No of Boxes in INCH	No of Boxes in CM	INCH / CM Client
1112	APPAREL	1372	0	Inch Client
1113	3PL	80	0	Inch Client
1114	3PL	39	1157	Cm Client
1115	CHEMICALS	1522	54	Inch Client
1116	MEDIA & PUBLICATIONS	5584	1	Inch Client
1117	APPAREL	1676	102	Inch Client
1118	INDUSTRIAL	80	0	Inch Client
1119	INDUSTRIAL	107	0	Inch Client
1120	APPAREL	318	0	Inch Client

Description:

Client ID	Ids of Clients
Industry Type	Industry type of the client
No. of Boxes in INCH	No. of boxes with dimensions in inch for the client
No. of Boxes in CM	No. of boxes with dimensions in centimeter for the client
INCH / CM Client	Dimension in which client makes orders the most

4. Sample data of analysis length for every client

Client ID	Max Length	Most Frequent Length	Frequency of Most Frequent Length	Frequency of Max Length	IQR - Lower Limit	IQR - Upper Limit
1113	58	24	19	1	20	24
1114	1235	130	439	1	35	130
1115	41	16	33	1	15	18
1116	118	18	677	1	18	18
1117	60	23	62	7	23	23.5
1118	120	57	4	1	44	57
1120	24	24	17	17	16	24

Description:

Client ID	Id for every client
Max Length	Maximum length for each client
Most Frequent Length	Most frequent length for each client
Frequency of Most Frequent Length	No, of times most frequent length occur for each client
Frequency of Max Length	No, of times maximum length occur for each client
IQR - Lower Limit	InterQuartieRange Lower for each client
IQR - Upper Limit	InterQuartieRange Upper for each client

5. Sample data of analysis weight for every client

Client ID	Max Weight	Most Frequent Weight	Frequency of Most Frequent Weight	Frequency of Max Weight	IQR - Lower Limit	IQR - Upper Limit
1113	432	20	6	1	22.5	96
1114	512	68	245	1	61	68
1115	4240	318	4	3	64.5	515.5
1116	4530	7	19	1	17	161
1117	1920	192	3	1	192	644
1118	112	50	2	1	50	66
1119	855.6	717.5	2	1	328.075	663.125

Description:

Client ID	Id for every client
Max Length	Maximum length for each client
Most Frequent Length	Most frequent length for each client
Frequency of Most Frequent Length	No, of times most frequent length occur for each client
Frequency of Max Length	No, of times maximum length occur for each client
IQR - Lower Limit	InterQuartieRange Lower for each client
IQR - Upper Limit	InterQuartieRange Upper for each client

6. Sample data after applying 5 kinds of logic

consignment_id	client_id	weight	volume	total_boxes	industry_type	CFT	length	breadth	height	unit	number_of_boxes	Logic_1_Outlier	Logic_2_Outlier	Logic_3_Outlier	Logic_4_Outlier	Logic_5_Outlier
32123	1112	13	3.23	1	APPAREL	4.024767	31	15	12	IN	1	FALSE	FALSE	FALSE	FALSE	FALSE
32124	1113	122	23.3	14	3PL	5.236051	22	12	8	IN	2	FALSE	FALSE	FALSE	FALSE	FALSE
32124	1113	122	23.3	14	3PL	5.236051	23	13	11	IN	4	FALSE	FALSE	FALSE	FALSE	FALSE
32124	1113	122	23.3	14	3PL	5.236051	20	13	11	IN	8	FALSE	FALSE	FALSE	FALSE	FALSE
32125	1112	19	3.23	1	APPAREL	5.882352	31	15	12	IN	1	FALSE	FALSE	FALSE	FALSE	FALSE
32126	1114	68	11.16	4	3PL	6.093189	132	122	12	CM	1	FALSE	TRUE	FALSE	FALSE	FALSE
32126	1114	68	11.16	4	3PL	6.093189	130	27	20	CM	1	FALSE	TRUE	FALSE	FALSE	FALSE

Description:

consignment_id	Every consignment id
client_id	Every client id
weight	Weight for every consignment
volume	Volume for every consignment
total_boxes	Total no. of boxes for every client
industry_type	Industry type for every client
CFT	Density for every consignment
length	Length for every consignment
breadth	Breadth for every consignment
height	Height for every consignment
unit	Unit of measurement for every consignment
number_of_boxes	Total no. of boxes for every consignment
Logic_1_Outlier	Outlier detected using Logic 1 for every consignment
Logic_2_Outlier	Outlier detected using Logic 2 for every consignment
Logic_3_Outlier	Outlier detected using Logic 3 for every consignment
Logic_4_Outlier	Outlier detected using Logic 4 for every consignment
Logic_5_Outlier	Outlier detected using Logic 5 for every consignment

7. Sample data after applying 5 kinds of logic with their performance

			No. of Error Detected					Model Performance %				
Client ID	Industry Type	No. of Consignments	Logic: 1	Logic: 2	Logic: 3	Logic: 4	Logic: 5	Logic: 1	Logic: 2	Logic: 3	Logic: 4	Logic: 5
1112	APPAREL	2598	2	3	9	0	0	0.07698229407	0.1154734411	0.3464203233	0	0
1113	3PL	385	2	2	7	0	0	0.5194805195	0.5194805195	1.818181818	0	0
1114	3PL	7811	7	904	6	356	0	0.0896172065	11.5734221	0.07681474843	4.557675074	0
1115	CHEMICALS	3668	20	0	13	15	0	0.5452562704	0	0.3544165758	0.4089422028	0
1116	MEDIA & PUBLICATIONS	10887	13	1	87	1	0	0.1194084688	0.009185266832	0.7991182144	0.009185266832	0
1117	APPAREL	5150	0	8	61	2	0	0	0.1553398058	1.184466019	0.03883495146	0
1118	INDUSTRIAL	185	0	13	0	0	0	0	7.027027027	0	0	0
1119	INDUSTRIAL	245	0	0	3	0	0	0	0	1.224489796	0	0
1194	INDUSTRIAL	10009	233	28	101	20	3	2.327904886	0.2797482266	1.009091817	0.1998201619	0.02997302428

Description:

Client ID	Ids of every client
Industry Type	Industry type for every client
No. of Consignments	Total no. of consignments for each client
No. of Error Detected using Logic 1	No. of error detected using Logic 1 for each client
No. of Error Detected using Logic 2	No. of error detected using Logic 1 for each client
No. of Error Detected using Logic 3	No. of error detected using Logic 1 for each client
No. of Error Detected using Logic 4	No. of error detected using Logic 1 for each client
No. of Error Detected using Logic 5	No. of error detected using Logic 1 for each client
Model Performance % using Logic 1	Performance of error detection using Logic 1 for each client
Model Performance % using Logic 2	Performance of error detection using Logic 2 for each client
Model Performance % using Logic 3	Performance of error detection using Logic 3 for each client
Model Performance % using Logic 4	Performance of error detection using Logic 4 for each client
Model Performance % using Logic 5	Performance of error detection using Logic 5 for each client

Analysis

1. For consignment_data:
    a. There are duplicates client_id but each with a new consignmet_id
    b. There are a total of 86663 entries with 11 features(or columns)
    c. The column of 'Volume' was named incorrectly
    d. Delivery_Date is the only column wiht missing(null) values
    e. There are a total to 12 unique industries and 1296 unique clients
    f. Industries with maximum number of client were as follows: [INDUSTRIAL > NON-ECOMMERCE > APPAREL > PHARMA > ELECTRONICS > MEDIA & PUBLICATIONS > FMCG > 3PL > RETAIL > ECOMMERCE > AUTOMOTIVE > CHEMICALS]
    g. There are some entries that contains 0 as volume which is impossible value
    h. Client 1151 has the maximum total weight, length and count for their consignments

2. For consignment_volume_data:
    a. Most measurements are done in INCH
    b. ‘created_at' column is useless as it contains no actual values
    c. There are a total of 86663 unique consignments
    d. Maximum number of boxes is for one consignment_id is 35300
    e. Maximum total no. of boxes is 5441 for one consignment_id
    f. Maximum no. of boxes in inches is for client: 1151
    g. Maximum no. of boxes in inches is for industry: ECOMMERCE
    h. Maximum no. of boxes in cm is for client: 1235
    i. Maximum no. of boxes in cm is for client: INDUSTRIAL
    j. No. of client using dimension is inches is higher that no. of clients using dimensions in cm
    k. There are duplicates consignment_id but each with unique values of length, breadth and height

3. For Outliers using 5 Logics:
    a. Logic 3 i.e. identifying outlier using IQR values for corresponding industry type has marked outliers the most, followed by Logic 2 then Logic 1 then Logic 4 and at last Logic 5 has marked outliers the least with count being just 3