REGION WISE HOSPITAL

DATA Analysis

Capstone Project

OBJECTIVE:

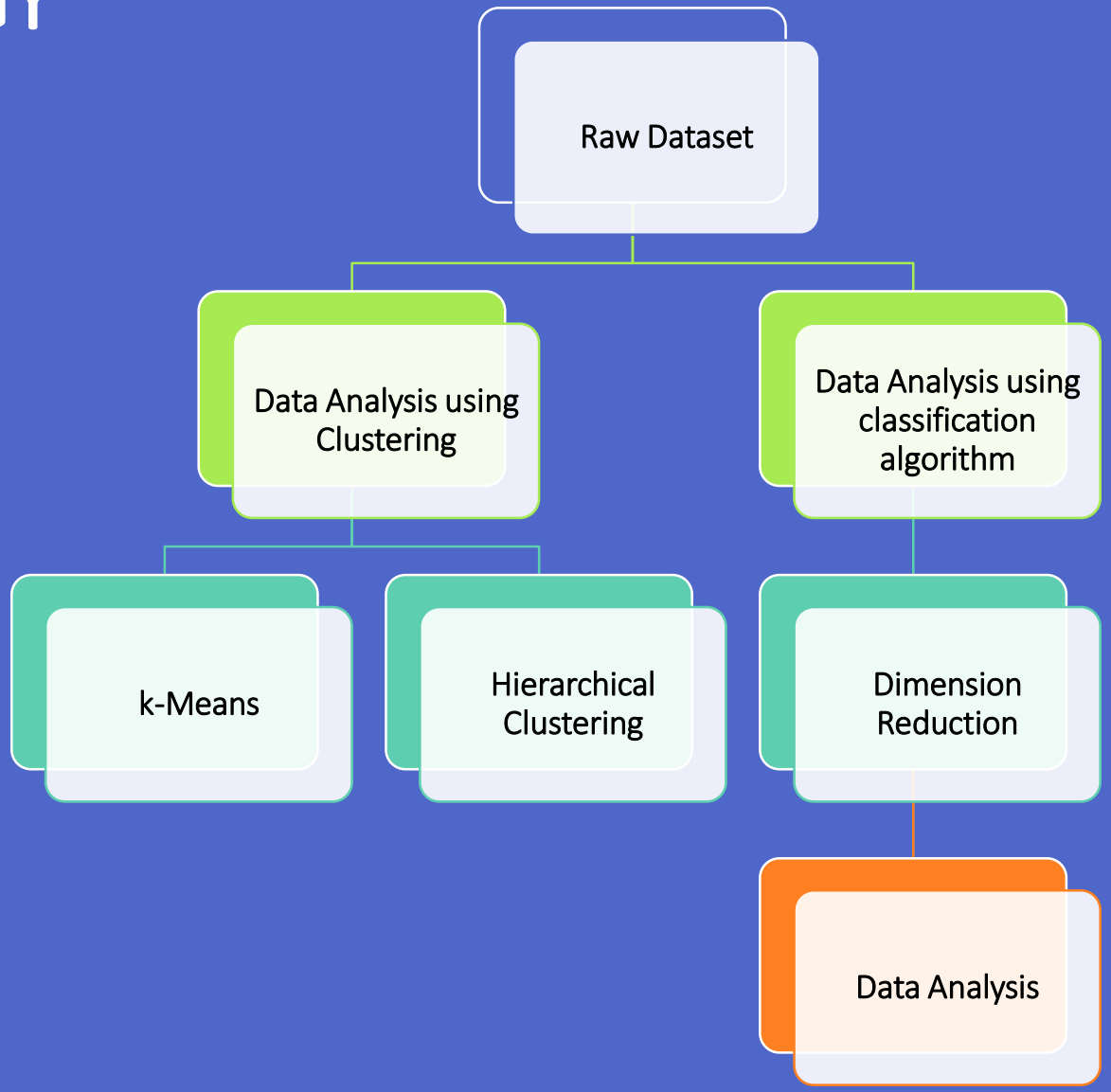
We have taken a dataset from hospital survey. We aim to :

- ❑ Analyze the data and draw conclusion on the length of stay of patients of a particular department
- ❑ Predict if below factors have any correlation with length of stay keeping in view the data accuracy.

DEPENDENT VARIABLES:

Region Code	Department	Severity of illness
Type of Admission	City Code	Deposit Amount
Age Group	Length of Stay	
Hospital Code	Hospital Type Code	

METHODOLOGY



DIMENSIONALITY REDUCTION

- Removal of not so significant dependent variables while retaining some meaningful properties of the original data.
- Less dimensions lead to less computation or training time.

INSIGNIFICANT PARAMETERS :

Case ID, Ward Type, Visitor with patient, Ward Facility Code, Available Extra Bed, Age

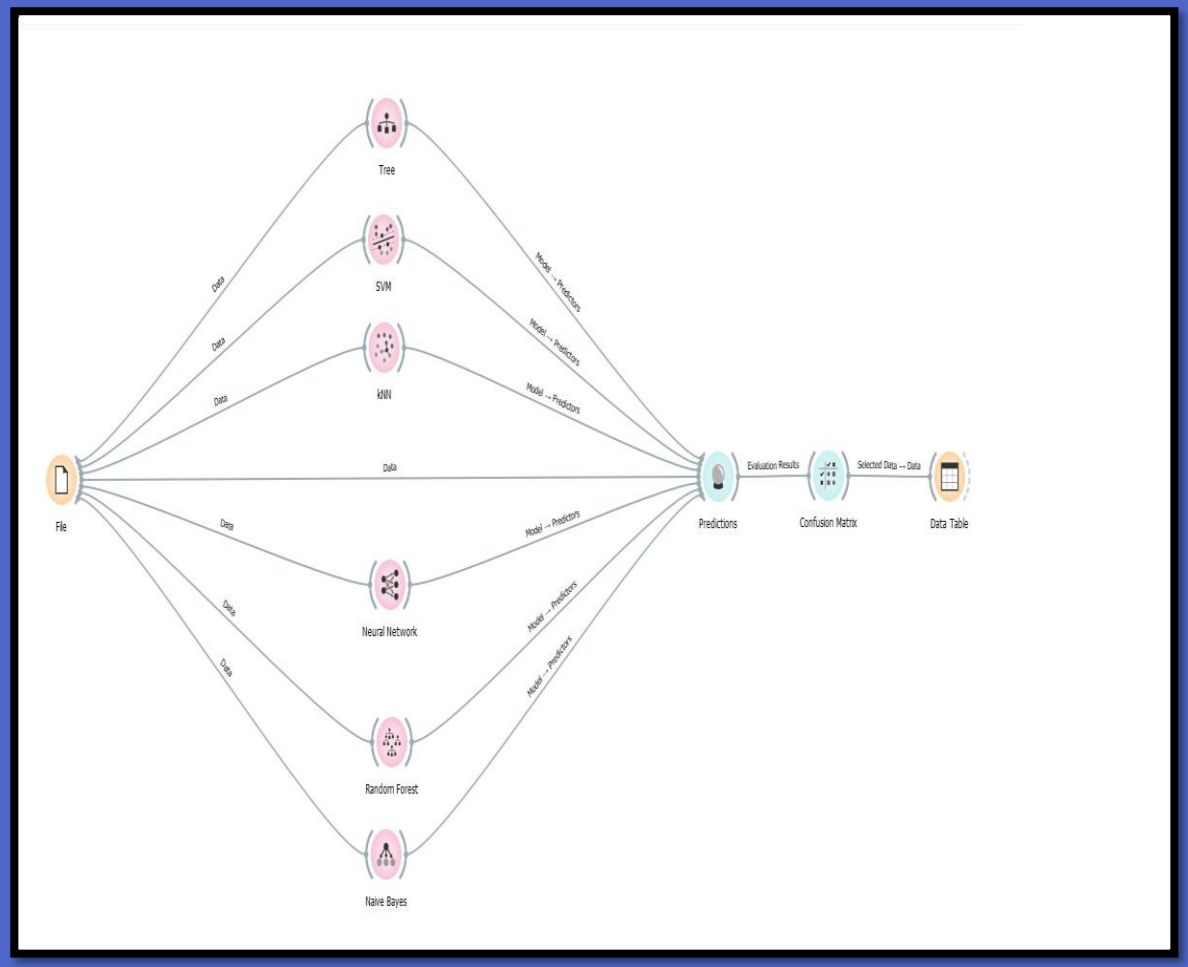
Column	Description
case_id	Case_ID registered in Hospital
Hospital_code	Unique code for the Hospital
Hospital_type_code	Unique code for the type of Hospital
City_Code_Hospital	City Code of the Hospital
Hospital_region_code	Region Code of the Hospital
Available Extra Rooms in Hospital	Number of Extra rooms available in the Hospital
Department	Department overlooking the case
Ward_Type	Code for the Ward type
Ward_Facility_Code	Code for the Ward Facility
Bed Grade	Condition of Bed in the Ward
patientid	Unique Patient Id
City_Code_Patient	City Code for the patient
Type of Admission	Admission Type registered by the Hospital
Severity of Illness	Severity of the illness recorded at the time of admission
Visitors with Patient	Number of Visitors with the patient
Age	Age of the patient
Admission_Deposit	Deposit at the Admission Time
Stay	Stay Days by the patient

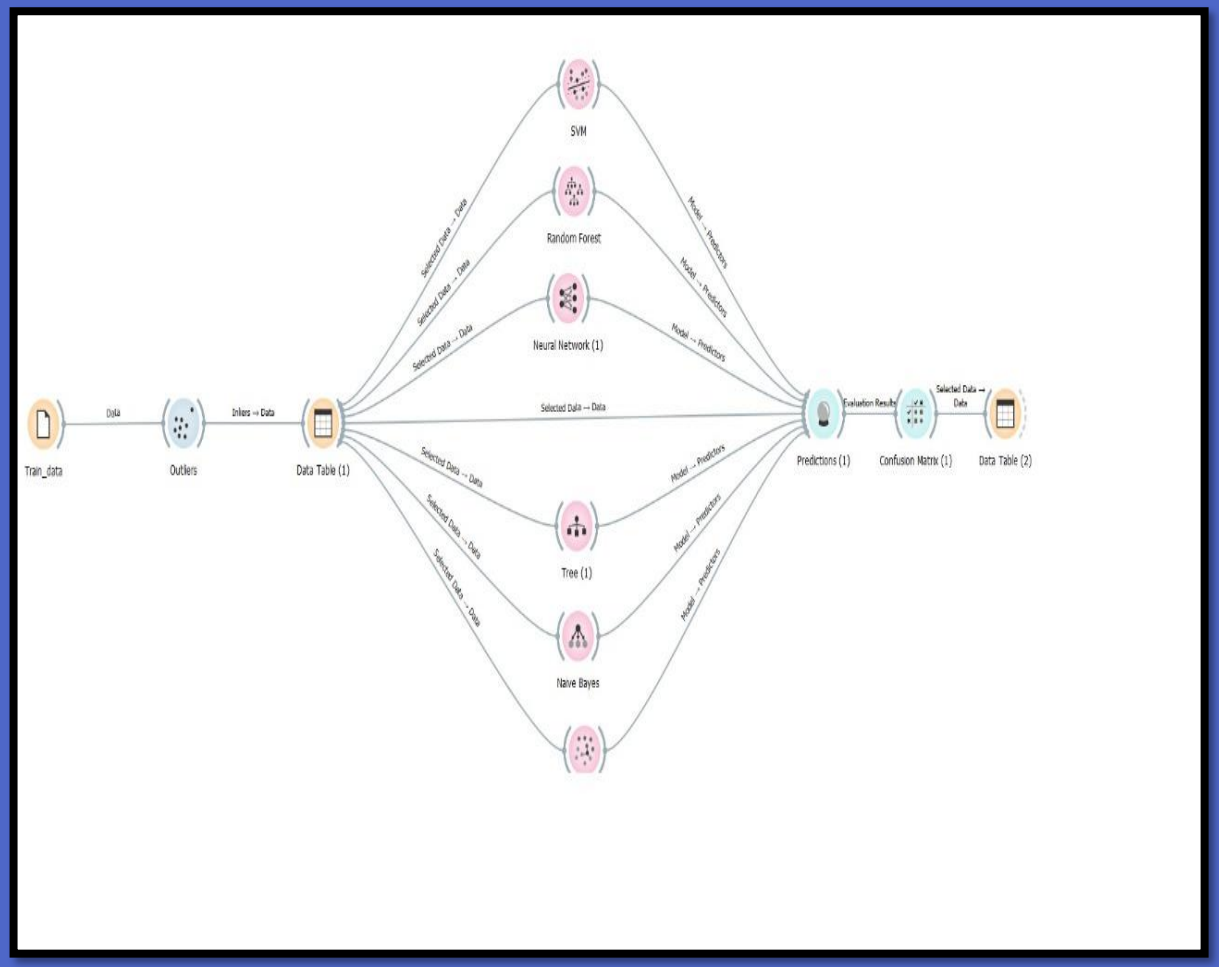
EXCEL ANALYSIS – To understand revenue aspects

Hospital_region_code	Department	Count of Patient	%	Extreme	Minor	Moderate	Extreme	Minor	Moderate	Average	Extreme	Minor	Moderate	Total	Incidence Age Group?
X	gynecology	101835	76%	15840	31051	54944	₹ 4,600	₹ 4,851	₹ 4,728	₹ 4,726	₹ 7,28,69,725	₹ 15,06,24,918	₹ 25,97,80,420	₹ 48,32,75,063	31-40
	anesthesia	15768	12%	2777	3926	9065	₹ 5,089	₹ 5,230	₹ 5,172	₹ 5,164	₹ 1,41,32,251	₹ 2,05,31,309	₹ 4,68,84,115	₹ 8,15,47,675	31-40
	radiotherapy	11000	8%	2083	2692	6225	₹ 4,680	₹ 5,032	₹ 4,889	₹ 4,867	₹ 97,49,337	₹ 1,35,45,959	₹ 3,04,32,514	₹ 5,37,27,810	41-50
	TB & Chest disease	4268	3%	756	1078	2434	₹ 4,930	₹ 5,205	₹ 5,044	₹ 5,060	₹ 37,27,088	₹ 56,11,398	₹ 1,22,76,783	₹ 2,16,15,269	51-60
	surgery	465	0%	127	47	291	₹ 4,875	₹ 4,936	₹ 4,916	₹ 4,909	₹ 6,19,170	₹ 2,31,998	₹ 14,30,460	₹ 22,81,628	51-60
X Total		**133336**		**21583**	**38794**	**72959**					**₹ 10,10,97,571**	**₹ 19,05,45,582**	**₹ 35,08,04,292**	₹ 64,24,47,445	
Y	gynecology	96874	79%	18778	24614	53482	₹ 4,637	₹ 4,918	₹ 4,788	₹ 4,781	₹ 8,70,75,297	₹ 12,10,56,843	₹ 25,60,46,896	₹ 46,41,79,036	31-40
	radiotherapy	12336	10%	2539	2684	7113	₹ 4,787	₹ 5,050	₹ 4,896	₹ 4,911	₹ 1,21,55,426	₹ 1,35,54,831	₹ 3,48,27,497	₹ 6,05,37,754	51-60
	anesthesia	9051	7%	1907	1904	5240	₹ 5,275	₹ 5,428	₹ 5,369	₹ 5,357	₹ 1,00,59,291	₹ 1,03,34,647	₹ 2,81,32,282	₹ 4,85,26,220	31-40
	TB & Chest disease	3582	3%	721	845	2016	₹ 5,079	₹ 5,230	₹ 5,188	₹ 5,166	₹ 36,62,229	₹ 44,19,622	₹ 1,04,58,902	₹ 1,85,40,753	51-60
	surgery	585	0%	267	39	279	₹ 4,896	₹ 4,904	₹ 4,898	₹ 4,899	₹ 13,07,214	₹ 1,91,241	₹ 13,66,602	₹ 28,65,057	51-60
Y Total		**122428**		**24212**	**30086**	**68130**					**₹ 11,42,59,457**	**₹ 14,95,57,184**	**₹ 33,08,32,179**	₹ 59,46,48,820	
Z	gynecology	50777	81%	8860	13841	28076	₹ 4,866	₹ 5,110	₹ 4,982	₹ 4,986	₹ 4,31,11,774	₹ 7,07,30,369	₹ 13,98,71,414	₹ 25,37,13,557	31-40
	radiotherapy	5180	8%	928	1418	2834	₹ 4,991	₹ 5,203	₹ 5,106	₹ 5,100	₹ 46,31,657	₹ 73,78,317	₹ 1,44,70,577	₹ 2,64,80,551	61-70
	anesthesia	4830	8%	769	1285	2776	₹ 5,567	₹ 5,621	₹ 5,575	₹ 5,588	₹ 42,81,106	₹ 72,22,896	₹ 1,54,76,463	₹ 2,69,80,465	41-50
	TB & Chest disease	1736	3%	325	436	975	₹ 5,193	₹ 5,364	₹ 5,262	₹ 5,273	₹ 16,87,792	₹ 23,38,861	₹ 51,30,449	₹ 91,57,102	61-70
	surgery	151	0%	35094	47066	102791	₹ 5,082	₹ 5,152	₹ 5,296	₹ 5,177	₹ 2,33,785	₹ 61,822	₹ 4,92,528	₹ 7,88,135	61-70
Z Total		**62674**		**45976**	**64046**	**137452**	₹ 25,699	₹ 26,451	₹ 26,221		**₹ 5,39,46,114**	**₹ 8,77,32,265**	**₹ 17,54,41,431**	₹ 31,71,19,810	
Grand Total		**318438**		**91771**	**132926**	**278541**					₹ 26,93,03,142	₹ 42,78,35,031	₹ 85,70,77,902		

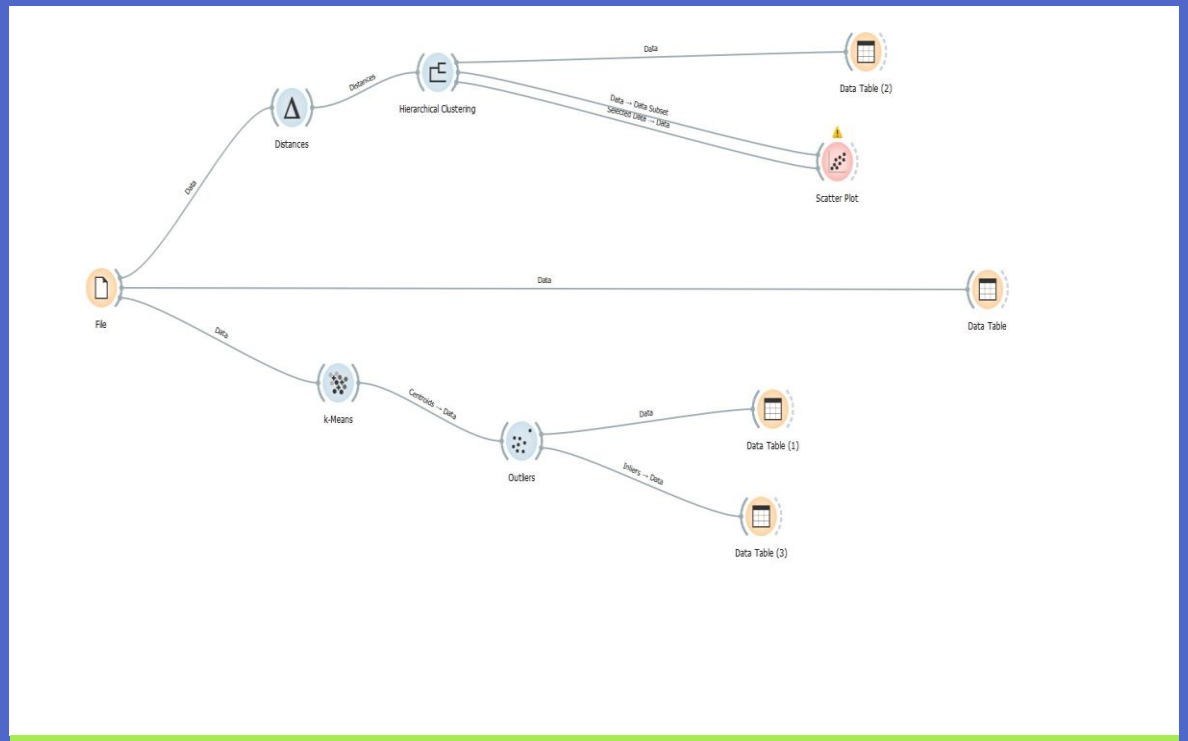
Excel Analysis – Basic Excel formula to understand the region wise top departments interms of volume and revenue
The Analysis reveled that the 3 regions have similar pattern

Dataset Analysis Model





DATA CLUSTERING



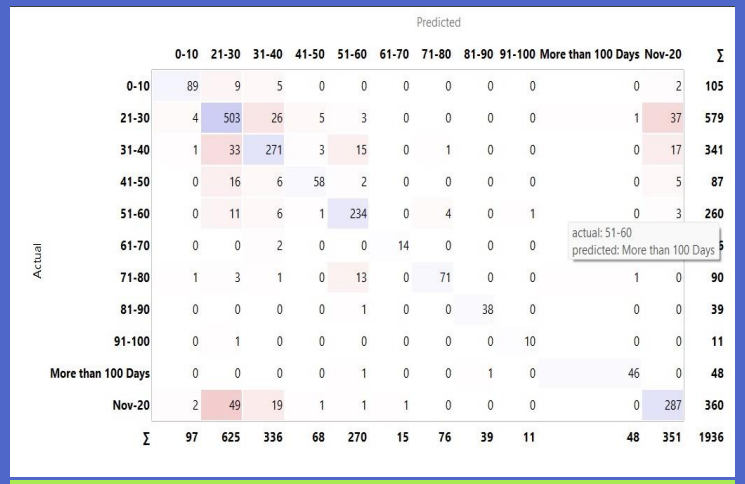
Clustering 1

	Stay	Cluster	Silhouette	Outlier	Hospital_code	lospital_type_cod	ity_Code_Hospita	ospital_region_coc	Department	Type of Admission	Severity of Illness
1	0-10	C2	0.547429	No	8	c	3	Z	radiotherapy	Emergency	Extreme
2	41-50	C2	0.585991	No	2	c	5	Z	radiotherapy	Trauma	Extreme
3	31-40	C4	0.571139	No	10	e	1	X	anesthesia	Trauma	Extreme
4	41-50	C3	0.584789	No	26	b	2	Y	radiotherapy	Trauma	Extreme
5	41-50	C3	0.584789	No	26	b	2	Y	radiotherapy	Trauma	Extreme
6	Nov-20	C1	0.528617	No	23	a	6	X	anesthesia	Trauma	Extreme
7	0-10	C1	0.553854	No	32	f	9	Y	radiotherapy	Emergency	Extreme
8	41-50	C1	0.52969	No	23	a	6	X	radiotherapy	Trauma	Extreme
9	51-60	C2	0.562136	No	1	d	10	Y	gynecology	Trauma	Extreme
10	31-40	C4	0.588695	No	10	e	1	X	gynecology	Trauma	Extreme
11	21-30	C1	0.550367	No	22	g	9	Y	radiotherapy	Urgent	Extreme
12	Nov-20	C3	0.576788	No	26	b	2	Y	radiotherapy	Urgent	Extreme
13	0-10	C2	0.505061	No	16	c	3	Z	radiotherapy	Emergency	Extreme
14	21-30	C2	0.565206	No	9	d	5	Z	radiotherapy	Urgent	Extreme
15	0-10	C2	0.530968	No	6	a	6	X	gynecology	Emergency	Extreme
16	0-10	C2	0.530968	No	6	a	6	X	gynecology	Emergency	Extreme
17	0-10	C1	0.530058	No	23	a	6	X	radiotherapy	Urgent	Extreme
18	Nov-20	C1	0.502824	No	29	a	4	X	anesthesia	Emergency	Extreme
19	21-30	C1	0.550713	No	32	f	9	Y	radiotherapy	Trauma	Extreme
20	31-40	C1	0.514115	No	12	a	9	Y	radiotherapy	Trauma	Extreme
21	31-40	C2	0.512956	No	16	c	3	Z	anesthesia	Trauma	Extreme
22	21-30	C2	0.565805	No	3	c	3	Z	anesthesia	Trauma	Extreme
23	Nov-20	C3	0.527029	No	21	c	3	Z	anesthesia	Trauma	Extreme
24	21-30	C2	0.53186	No	6	a	6	X	anesthesia	Urgent	Extreme

Clustering 2

We made 4 clusters C1, C2, C3 & C4 – Depending on the Silhouette Score recommendation
We explored 2 methods 1. Hierarchical Clustering , 2. k-Means

DATA ANALYSIS



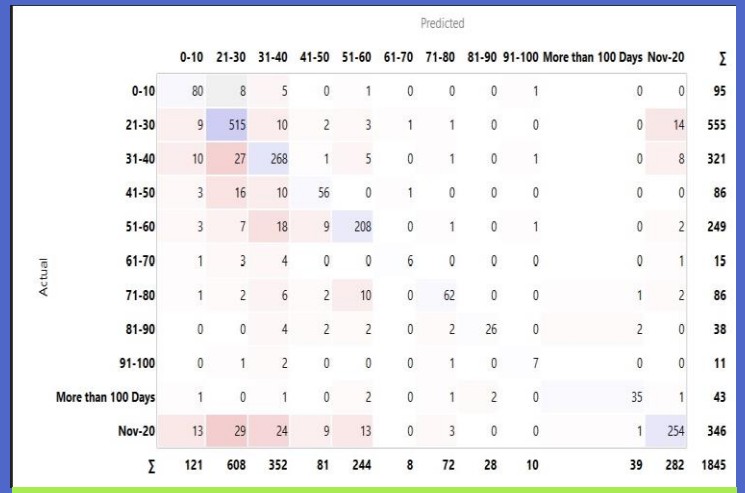
Actual \ Predicted	0-10	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	More than 100 Days	Nov-20	Σ
0-10	89	9	5	0	0	0	0	0	0	0	2	105
21-30	4	503	26	5	3	0	0	0	0	1	37	579
31-40	1	33	271	3	15	0	1	0	0	0	17	341
41-50	0	16	6	58	2	0	0	0	0	0	5	87
51-60	0	11	6	1	234	0	4	0	1	0	3	260
61-70	0	0	2	0	0	14	0	0	0			
71-80	1	3	1	0	13	0	71	0	0	1	0	90
81-90	0	0	0	0	1	0	0	38	0	0	0	39
91-100	0	1	0	0	0	0	0	0	10	0	0	11
More than 100 Days	0	0	0	0	1	0	0	1	0	46	0	48
Nov-20	2	49	19	1	1	1	0	0	0	0	287	360
Σ	97	625	336	68	270	15	76	39	11	48	351	1936

actual: 51-60
predicted: More than 100 Days

Raw Confusion Matrix

Actual \ Predicted	0-10	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	More than 100 Days	Nov-20	Σ
0-10	31	37	13	0	14	0	1	0	0	0	10	106
21-30	21	370	86	2	53	0	1	0	0	8	45	586
31-40	8	139	137	2	33	0	1	0	0	3	29	352
41-50	3	32	34	5	10	0	0	0	0	2	4	90
51-60	11	97	54	2	82	0	0	0	0	3	27	276
61-70	2	3	9	0	2	0	0	0	0	0	0	16
71-80	3	45	19	0	12	0	4	0	0	2	11	96
81-90	1	19	6	0	7	0	0	0	0	1	6	40
91-100	0	5	4	0	0	0	0	0	0	1	1	11
More than 100 Days	3	20	9	0	3	0	0	0	0	13	3	51
Nov-20	17	176	56	1	31	0	3	0	0	6	86	376
Σ	100	943	427	12	247	0	10	0	0	39	222	2000

Dimensional Reduction Confusion Matrix



Actual \ Predicted	0-10	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	More than 100 Days	Nov-20	Σ
0-10	80	8	5	0	1	0	0	0	1	0	0	95
21-30	9	515	10	2	3	1	1	0	0	0	14	555
31-40	10	27	268	1	5	0	1	0	1	0	8	321
41-50	3	16	10	56	0	1	0	0	0	0	0	86
51-60	3	7	18	9	208	0	1	0	1	0	2	249
61-70	1	3	4	0	0	6	0	0	0	0	1	15
71-80	1	2	6	2	10	0	62	0	0	1	2	86
81-90	0	0	4	2	2	0	2	26	0	2	0	38
91-100	0	1	2	0	0	0	1	0	7	0	0	11
More than 100 Days	1	0	1	0	2	0	1	2	0	35	1	43
Nov-20	13	29	24	9	13	0	3	0	0	1	254	346
Σ	121	608	352	81	244	8	72	28	10	39	282	1845

Outlier Reduction Confusion Matrix

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.974	0.837	0.837	0.839	0.837
kNN	0.859	0.486	0.462	0.479	0.486
Naive Bayes	0.745	0.335	0.351	0.419	0.335
SVM	0.865	0.488	0.464	0.525	0.488
Random Forest	0.994	0.905	0.904	0.910	0.905
Tree	0.988	0.829	0.827	0.832	0.829

RAW DATA Prediction

Model	AUC	CA	F1	Precision	Recall
kNN	0.710	0.364	0.332	0.353	0.364
Neural Network	0.758	0.385	0.339	0.407	0.385
Random Forest	0.823	0.457	0.432	0.477	0.457
SVM	0.640	0.310	0.217	0.355	0.310
Tree	0.824	0.432	0.398	0.449	0.432
Naive Bayes	0.622	0.282	0.251	0.273	0.282

Dimensional Reduction Prediction

Model	AUC	CA	F1	Precision	Recall
Tree (1)	0.987	0.822	0.821	0.828	0.822
Neural Network	0.973	0.827	0.826	0.830	0.827
SVM	0.874	0.478	0.448	0.547	0.478
Naive Bayes	0.742	0.332	0.345	0.415	0.332
kNN	0.858	0.483	0.459	0.476	0.483
Random Forest	0.994	0.909	0.909	0.915	0.909

Outlier Reduction Prediction

INFERENCES

- Excel Analysis revealed that age group trend across Gynecology across 3 regions is similar, and except region Z other locations showed same trend for other specialties as well
- Tree Analysis resulted in best accuracy in case of raw data analysis and dimensional reduction analysis
- Random Forest resulted in best accuracy results in Outlier reduction analysis
- Dimensional Reduction has led to Accuracy Reduction
- Outlier Reduction has led to Accuracy Improvement
- Due to Data Over-fitting, Raw data prediction came out better than Dimensional & Outlier Reduction
- The model is able to predict the length of stay with an accuracy of 99.4%