Final Report:
Using Patient Triage and Historical Data to Predict the disposition before seen by the provider

Problem statement:

According to CDC, the national number of Emergency Department visit is 130 million in year 2018. Out of those visits, there were about 12.4% (16.2 million) resulting in hospital admission. Many times, patients who has long ED length of stay were waiting the bed to be available in Hospital Inpatient units. To help the hospital providing adequate resources and shorten the ED Length of Stay for a patient, it will be very helpful if there is a way to predict hospital admissions right after Triage.

Can use machine learning to help predict hospital admission at the time of ED triage using patient history in addition to information collected at triage?

Data Wrangling:

The raw dataset from Woo Suk Hon, Adrian Daniel Haimovich and R.Andrw Tayor has 972 variables with total of 560,486 adult patients. Because it's a very huge dataset, so there is a need to do a little cleaning.

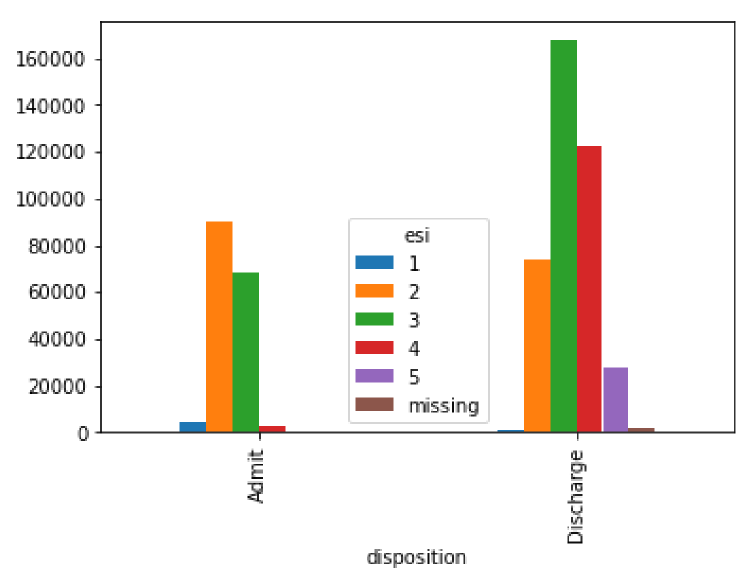
First, find out what are the columns with NaN data and the need to fill the NaN with propriate value. For Binary Type, fill 0 for all the missing values, For Categorical type, fill the missing according the the category, most are other or missing or refused.

I didn't delete any row, so the final shape of the dataset is still 972 columns with 560486 rows.

Exploratory Data Analysis:

To understand the data, I did counts to understand:

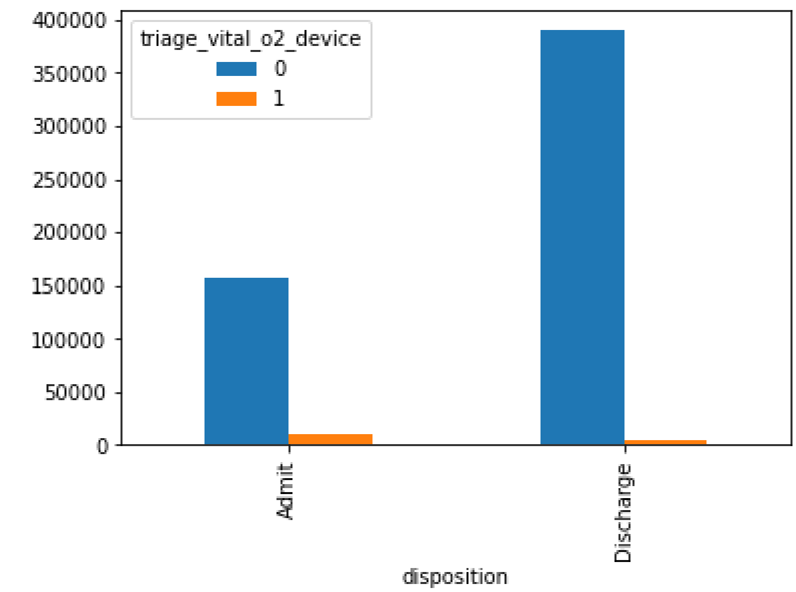
- how many patients were admitted and how many were Direct Discharge from ED.
- The counts for ESI, which is the Acuity Level
- Compare the number patients of different ESI from Admit to Discharge



The Graph shows that patients who were with ESI 1 and 2 were Admit from ED more then patient who where with ESI 4 and 5.

The majority of ESI 3 patients also direct discharge from the ED as well

- Compare the number of patient with O2_decive when entering ED from Admit to Discharge



Patient came in with O2 device did got admit to the hospital more.

- The count of chief-complains when patient coming to ED and compare it’s disposition

	cc_abdominalcramping	cc_abdominaldistention	cc_abdominalpain	cc_abdominalpainpregnant	cc_abnormallab	cc_abscess	cc_addictionproblem	cc_agitation
disposition								
Admit	70	383	19486	152	2716	446	92	286
Discharge	514	244	34844	1902	1010	3068	607	368

cc_alcoholintoxication	cc_alcoholproblem	cc_allergicreaction	cc_alteredmentalstatus	cc_animalbite	cc_ankleinjury	cc_anklepain	cc_anxiety	cc_arminjury	cc_armpain
1901	434	346	6563	151	153	226	496	51	569
14034	1804	2979	1989	1491	1948	2704	2407	717	3304

cc_armswelling	cc_assaultvictim	cc_asthma	cc_backpain	cc_bleeding/bruising	cc_blurredvision	cc_bodyfluidexposure	cc_breastpain	cc_breathingdifficulty	cc_breathingproblem
221	283	411	3063	234	141	0	37	2067	194
501	2947	1917	17575	616	370	817	480	1209	251

cc_burn	cc_cardiacarrest	cc_cellulitis	cc_chestpain	cc_chesttightness	cc_chills	cc_coldlikesymptoms	cc_confusion	cc_conjunctivitis	cc_constipation	cc_cough	cc_cyst
20	295	84	16071	320	181	317	548	5	449	2568	66
661	173	121	19717	814	416	2868	239	431	1354	9986	1119

cc_decreasedbloodsugar-symptomatic	cc_dehydration	cc_dentalpain	cc_depression	cc_detoxevaluation	cc_diarrhea	cc_dizziness	cc_drug/alcoholassessment	cc_drugproblem	cc_dyspnea
650	368	45	593	341	1620	4381	163	394	2018
507	303	3771	1230	884	2725	8404	1102	2325	1042

cc_dysuria	cc_earpain	cc_earproblem	cc_edema	cc_elbowpain	cc_elevatedbloodsugar-nosymptoms	cc_elevatedbloodsugar-symptomatic	cc_emesis	cc_epigastricpain	cc_epistaxis	cc_exposuretostd
244	80	15	471	54	569	917	4838	325	315	1
1425	3105	428	362	576	822	736	6675	598	1658	301

cc_extremitylaceration	cc_extremityweakness	cc_eyeinjury	cc_eyepain	cc_eyeproblem	cc_eyeredness	cc_facialinjury	cc_faciallaceration	cc_facialpain	cc_facialswelling	cc_fall
47	529	52	78	207	11	62	18	67	295	5643
2083	307	640	1700	2245	689	600	650	839	1294	13376

cc_fall>65	cc_fatigue	cc_femaleguproblem	cc_fever	cc_fever-75yearsorolder	cc_fever-9weeksto74years	cc_feverimmunocompromised	cc_fingerinjury	cc_fingerpain	cc_fingerswelling	cc_flankpain
3335	2291	466	190	1075	3060	505	39	62	59	1955
4002	1621	2718	333	212	3092	112	2275	1117	512	6972

I can’t list all the tables due to too many chief-complains, however, the take does show that some chief complains has higher number of patients been admit to hospital.

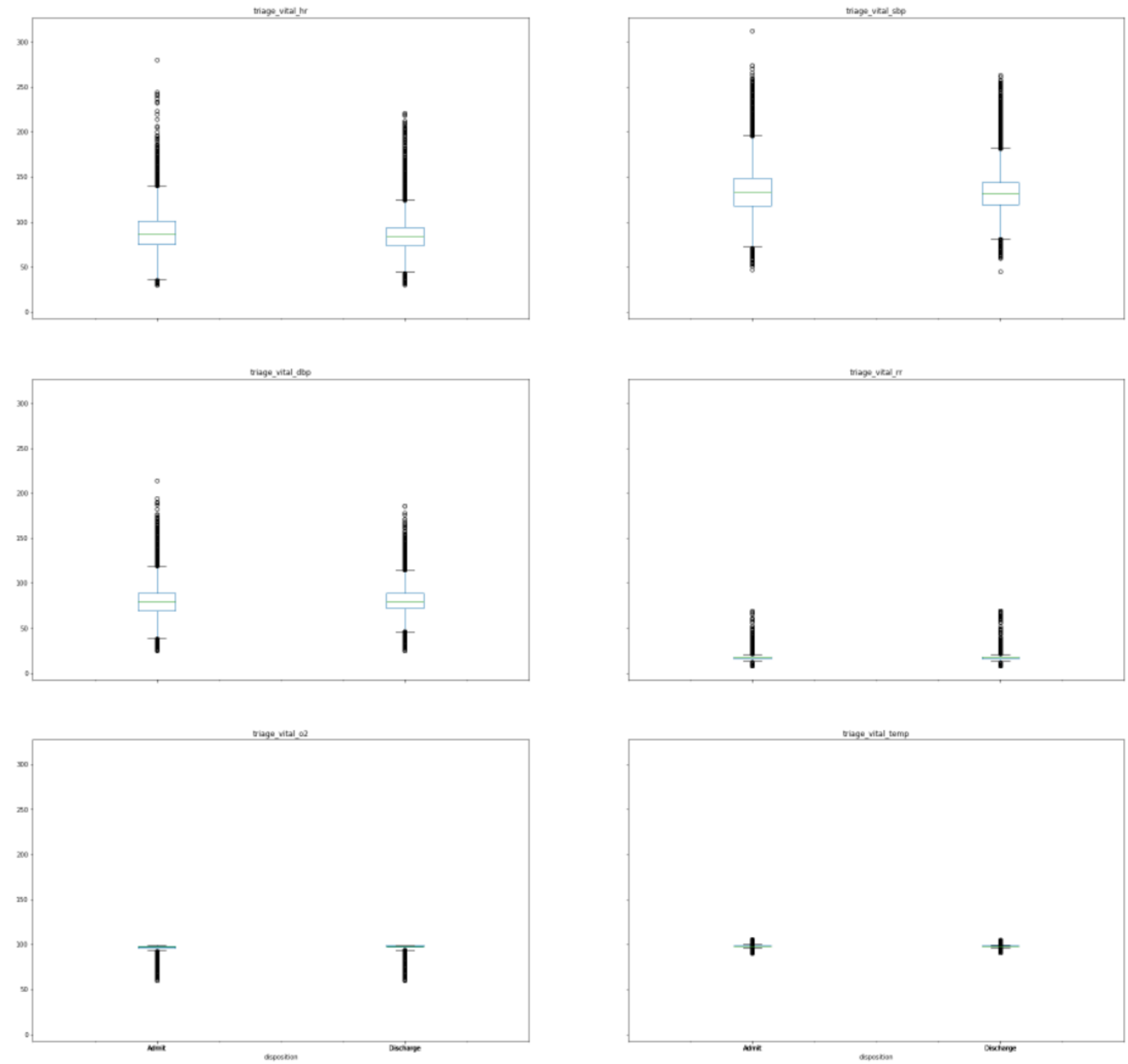
I also use describe function to understand the summary of Triage Only data for Admit patients and discharge patient. To see if there is large difference between the mean, std, and others. Also create a box plot to see the difference between each Triage Vital and it’s disposition.

```
AdmitTriVital.describe()
```

	triage_vital_hr	triage_vital_sbp	triage_vital_dbp	triage_vital_rr	triage_vital_o2	triage_vital_temp
count	94841.000000	94206.000000	94148.000000	93412.000000	73586.000000	87809.000000
mean	88.862366	134.742079	79.408987	17.980238	96.639427	98.159803
std	19.532877	24.834481	16.158699	2.428710	2.624910	0.966250
min	30.000000	47.000000	25.000000	8.000000	60.000000	90.000000
25%	75.000000	118.000000	69.000000	16.000000	96.000000	97.600000
50%	87.000000	133.000000	79.000000	18.000000	97.000000	98.100000
75%	101.000000	149.000000	89.000000	18.000000	98.000000	98.500000
max	280.000000	312.000000	214.000000	69.000000	99.000000	106.000000

```
DCTriVital.describe()
```

	triage_vital_hr	triage_vital_sbp	triage_vital_dbp	triage_vital_rr	triage_vital_o2	triage_vital_temp
count	299593.000000	298376.000000	298275.000000	296587.000000	215583.000000	289692.000000
mean	84.558468	132.930294	80.826007	17.502648	97.480767	98.061215
std	15.579172	20.062015	13.042584	1.765464	1.575886	0.717601
min	30.000000	45.000000	25.000000	8.000000	60.000000	91.000000
25%	74.000000	119.000000	72.000000	16.000000	97.000000	97.700000
50%	84.000000	131.000000	80.000000	18.000000	98.000000	98.000000
75%	94.000000	144.000000	89.000000	18.000000	99.000000	98.400000
max	221.000000	263.000000	186.000000	69.000000	99.000000	105.200000



The median of triage vitals shows not much difference from Admit to Discharge. However, the range of the vital is larger in Admit patients.

Model Selection:

I split the date into three different datasets: Triage Data Only, Historical Data Only and the full dataset. I want to see if it will make difference and if we can reduce the variables to do the prediction.

I test two different machine learning classification models: Logistic Regression, GradientBoosting. The Metric I focused on is the Accuracy. I want my model to correctly predict the disposition.

Before building the models, I deleted few variables that I think wont have much impact, like religion, employstatus, insurance_status….etc. Due to the nature of the machine learning, I also can't leave any null values in the dataset, so I search the best to way to fill the null value with extreme values (-9999). According to the experts, the machine learning model will ignore the extreme values.

When it comes to the models, I use mostly the default value. However, in Logistic Regression, due to the data size, I need to set a larger number of max_iter in order for it to work. I did a GridSearch Cross Validation, the model is doing will with default value. Compare to three different datasets. The full dataset performed the best in terms of predicting the correct disposition, it's has both high Accuracy score and High AUC Score.

For GradientBoosting, I use most the default value as well. However, I did try to tun the best learning rate. With three different datasets. Only the Full dataset had the different learning rate. Also, from the result. The Full dataset performed the best in terms of predicting the correct disposition.

Both two models are all work well with he prediction, however, GradientBoosting is performing a little bit better with higher Accuracy score and AUC score.

In Conclusion:

To be able to predict correctly, we can't just use partial information to do the determination. We need to include all the Triage and Historical information from the patient in order to predict correctly. The features are important for the machine to learn.