Predicting Patient Length of Stay via Machine Learning and Correlation Analysis: A Data-Driven Approach to Optimize Hospital Resource Allocation and Operational Efficiency

Project implementation - Objective 1 & 2

In this document, we will discuss our Objective 1 & 2, the methods we used to perform our analysis, and the results we achieved, with proper evaluation. Starting from stating our aim, objectives 1 and 2 aimed to achieve the following:

Objective 1
Identification of all variables that present a positive correlation with patients' length of stay: our first goal is study the correlation between all dataset variables and LOS, extrapolating the relevant ones and giving an explanation to why they present a positive or negative correlation.

Objective 2
Identification of clusters of patients whose length of stay substantially deviates from the average: after studying all relevant correlations, we will analyze cross-correlations and try to cluster patients whose characteristics make them more likely to stay hospitalized for a prolonged amount of time.

The following discussion will cover the methods used during our analysis for both objectives.

Methods

Data acquisition and preparation
To handle and prepare our data, we used Python. After importing the data into Python, we first preprocessed the data. We checked whether there were missing values in our data, and got the result observable in *Resource 1* (Resources and Tables section).

To handle the missing values for Bed Grade and City_Code_Patient, we used the median to fill in the Bed Grade column and the mode to fill in the City_Code_Patient column. After processing the missing values, we encoded the target variable. Since the target variable 'Stay' is a binned categorical variable, we encoded it from low to high according to the bin position. Then, we got a new variable - 'Stay number'.

For objective 2, some of our categorical variables have been transformed into dummy variables, which have been again transformed into numerical variables for proper handling.

Following is a detailed description of the methods used for Objective 1 and Objective 2, sequentially. All resources and tables can be found at the end of the document, in the *Resources and Tables* section.

Methods for Objective 1

Correlation between categorical variables and LOS
In this article, categorical variables refer to variables that have no natural order between different categories, such as 'Department'. The categorical variables include:
'Hospital_code', 'Hospital_type_code', 'City_Code_Hospital', 'Hospital_region_code', 'Department', 'Ward_Type', 'Ward_Facility_Code', 'City_Code_Patient', 'Type of Admission',

In this article, we first used the one-hot method to process each categorical variable by generating dummy variables. The generated dummy variables were then used as X variables, and the target variable 'Stay_number' has been used as Y (dependent variable) for OLS regression. The resulting output obtained through this method is shown in *Resource 2*.

With this method, it is evident that the regression results are very poor, the R square value is only 0.073, and there are serious multicollinearity problems; therefore, this method has been deemed not sufficiently explanatory.

For this reason, we chose another method to calculate the correlation ratio of each variable to the target variable (https://en.wikipedia.org/wiki/Correlation_ratio). The value of this ratio ranges from 0 to 1. The value 0 means that the categorical variable has almost no difference/no correlation with the numerical variable; 1 means that the categorical variable completely distinguishes the numerical variable (extreme case); the value between 0 and 1 indicates the degree of explanation of the categorical variable to the numerical variable. We got the following results:

Hospital_code vs Stay_number η: 0.15
Hospital_type_code vs Stay_number η: 0.08
City_Code_Hospital vs Stay_number η: 0.09
Hospital_region_code vs Stay_number η: 0.01
Department vs Stay_number η: 0.04
Ward_Type vs Stay_number η: 0.19
Ward_Facility_Code vs Stay_number η: 0.08
City_Code_Patient vs Stay_number η: 0.08
Type of Admission vs Stay_number η: 0.09

The correlation ratio between each categorical variable and the target variable can be obtained by calculation ratio, and the correlation between each variable and LOS can be further measured.

Correlation between ordinal variables and LOS
After calculating the correlation between categorical variables and LOS, this paper will next study the correlation between ordinal variables and LOS. Ordinal variables include continuous numerical variables, such as ‘Admission_Deposit’, and categorical variables with natural ordering of different categories, such as ‘Age’. For continuous numerical variables, they can be used directly as long as the integrity of the data is guaranteed; for ordered categorical variables, we first encoded them from small to large according to

the degree of classification. There are two ordered categorical variables: ‘Severity of Illness’ and ‘Age’. We encoded two new variables: 'Severity_number' and ‘Age_number’.

After encoding, we calculated the Spearman correlation coefficient matrix between the ordered variables and the target variable; we then drew a heat map (*Table 1*). At the same time, we calculated the Spearman correlation coefficient between each ordered variable and the target variable separately, and visualized their relationship (*Table 2*).
The ordered variables involved in the calculation are: 'Available Extra Rooms in Hospital', 'Bed Grade', 'Severity_number', 'Age_number', 'Visitors with Patient', 'Admission_Deposit'. The target variable is ‘Stay_number’. We calculated the following results:

Available Extra Rooms in Hospital Spearman correlation coefficient: -0.11
p-value: 0.0
Bed Grade Spearman correlation coefficient: -0.0018
p-value: 0.31
Severity_number Spearman correlation coefficient: 0.13
p-value: 0.0
Age_number Spearman correlation coefficient: 0.09
p-value: 0.0
Visitors with Patient Spearman correlation coefficient: 0.42
p-value: 0.0
Admission_Deposit Spearman correlation coefficient: -0.06
p-value: 0.0

Methods for Objective 2

For Objective 2, we wanted to individuate clusters of patients whose specific features would stand out in relation to their length of stay. To do so, we used R and Radiant visualization tools to create cross distribution plots of our variables of interest.

First of all, we used R to install the Radiant package through the command: install.packages("radiant", repos = "https://radiant-rstats.github.io/minicran/", type = 'binary'). Secondly, we launched the Radiant software through the R command: radiant::radiant(); radiant::radiant(browser = TRUE). Once in the software, in the data management window, we uploaded our dataset in csv format (comma-separator, period-decimal) in its entirety (318,438 observations). We then moved the dataset to the data visualization section, where we set “plot-type” to cross-distribution. Here, we selected a combination of X-variables of interest, and selected the “X-variables combination” option to put each in a proportional relation with the other. The “Length of stay” variable (binned into ten 10-days groups, plus an 11th category for any length of stay above 100 days) has been set as the facet row variable for all our plots, showing on the right Y-axis for each bin. The facet column variable has instead been picked depending on the X-variables of interest. Valuable results have been found across three different arrangements of our variables:

Table (3) plots the cross distribution of X-variables “Available rooms” and “Visitors with patients” in relation to “Length of stay” (facet row) and “Department” (facet column).

Table (4) establishes a relation between our dummy variables for grade of “Admission” and “Severity of illness”, “Length of stay” (facet row), and “Type of admission” (facet column).

Tavle (5) presents “Visitors with patient” as the X-variable, put in relation to “Length of stay” (facet row) and “Age” (facet column).

The outcome of the cross-distribution plots analysis will be discussed in the *Results* section.

Software implementation
In Objective 1, we used Python. For Objective 2, we used R and Radiant.

Results

Objective 1
First, for objective 1, we mainly studied the correlation between each variable and the target variable, that is, the correlation analysis between each variable and the patient's stay time.

We divided independent variables into two categories, unordered variables and ordered variables, where unordered variables refer to ordinary categorical variables, and ordered variables refer to continuous numerical variables and ordered categorical variables. For the first category of variables, we used One-Hot and regression, as well as the method of calculating the correlation ratio to calculate the correlation. For the second category of variables, we calculated the Spearman correlation coefficient for measurement; the specific steps have been explained in the method part.

Our Findings

After analyzing the correlation between the two types of variables and the target variable, we drew the following conclusions:

1. When the common categorical variables are processed one-hot to generate dummies variables and then OLS regression is performed with the target variable ‘Stay_number’, the regression results are poor and there is a serious multicollinearity problem, which to some extent shows that the common categorical variables have no strong correlation with LOS.

2. First, among the common categorical variables, there are certain differences in ‘Stay_number’ in different groups of ‘Hospital_code’ and ‘Ward_Type’, indicating that the two have a certain impact on the length of stay of patients in the hospital; while the remaining variables such as: ‘Hospital_type_code’, ‘City_Code_Hospital’, ‘Hospital_region_code’, ‘Department’, ‘Ward_Facility_Code ’, ‘City_Code_Patient ’, and ‘Type of Admission’ have no significant impact on the length of stay of patients in the hospital.

3. For the three variables ‘Available Extra Rooms’, ‘Bed Grade’, and ‘Admission_Deposit’ in the ordered categorical variables, the calculated Spearman correlation coefficient values and the corresponding p-values, as well as *Table 1*, show that the three have a negative correlation with LOS. Except for the p-value of ‘Bed Grade’ of 0.31, which is not statistically significant, the other two correlation coefficients are statistically significant in reducing the patient’s length of stay.

4. For the variables ‘Severity_number’, ‘Age_number’, and ‘Visitors with Patient’, according to the calculated spearman coefficients and *Table 1*, it can be clearly seen that they are positively correlated with LOS. And since the calculated p-values are all 0, the increase in the three variables themselves statistically significantly prolongs the length of hospital stay of patients.

This result will greatly help us to carry out the next step of our research, as we need to use machine learning to build a model that can be used to predict LOS, and the results of correlation analysis can help us figure out the general prediction direction to a certain extent. At the same time, the results of correlation analysis can also help us to select appropriate independent variables for model construction in machine learning to a certain extent, which is very helpful for enhancing the accuracy of the model.

In addition, by studying the correlation between variables and hospital stay, we can help hospitals to infer the approximate hospital stay when they know the basic information about patients, helping to improve management efficiency. As an example, if many visitors are with the patient, the patient's hospital stay may be longer, and vice versa. Then, the hospital can reasonably arrange beds based on the results, and improve medical management efficiency.

Objective 2

As previously stated, for our Objective 2 we wanted to individuate clusters of patients whose specific features would stand out in relation to their length of stay: in order to explore such relationships, we wanted to analyze the cross-distributions and cluster patients whose characteristics make them more likely to stay hospitalized for a prolonged amount of time.

To do so, we used R and Radiant visualization tools to create cross-distribution plots of our variables of interest, the functioning of which has been explained in the *Methods* section. These plots visually show potential clusters that present a prolonged length of stay in comparison to all the associated sub-categories.

Our findings

After a careful analysis of our results, multiple clusters of patients have been highlighted. From table (3), we can see a neat difference in patients of the gynecology department. Within these patients, the ones to present a longer length of stay also present a higher number of visitors with them, and are usually allocated in hospitals with an abundance of extra rooms. This suggests that gynecology patients with multiple visitors and in hospitals not in a rush to free up rooms are more likely to extend their stay under the hospital’s cures, potentially because of gynecologic diseases’ need for extended treatments.

In table (4), it is important to firstly state that the "Admission_transformed" variable assumes values of Trauma (1), Urgent (2), Emergency (3), while the "Severity_of_illness" variable assumes Minor (1), Moderate (2), Extreme (3). In light of this information, it is possible to notice that our data clusters around emergency patients with moderate or extreme degrees of illness, and Trauma patients with minor or moderate degrees of illness. A minor cluster can also be observed for urgent patients at moderate risk.

For table (5), the number of visitors for each patient is put in cross-relation to their age and length of stay. The visual analysis evidently shows a cluster around the central-bottom area of the plot, which is represented by patients between 30-80 years old who stay more than 80 days hospitalized, and have an average higher number of visitors.

These results suggest important insight for the next steps of our research. First of all, gynecology patients are more likely to be hospitalized, and their length of stay is expected to be longer if they receive multiple visitors. Another important insight is noticing how if the hospital has an abundance of spare rooms, it won't be in a rush of discharging patients, and their length of hospitalization is expected to be prolonged.

Further in the analysis, we found significance in the results of table (9) as they highlight clusters for each type of admission that are more likely to be hospitalized and have a prolonged stay. Thanks to our results, we can rule out patients with minor emergencies, extreme traumas, and minor or extreme urgencies from the categories of patients at risk of prolonged hospitalization. Instead, we can focus on the complimentary categories, especially relatively higher degrees of emergencies, and lower degrees of traumas.

Finally, key insights also come from our last plot analysis, the result of which has been discussed above. Saying that patients who stay hospitalized for more than 80 days have, on average, a higher number of visitors and are between 30-80 years old, let us rule out a huge portion of the population from our risk of prolonged stay analysis. A minor assumption that can be made is that patients with more visitors are possibly affected by non-lethal conditions which however required long hospitalizations, justifying the number of relatives and friends coming to visit. On the other hand, a major assumption that can be made is that younger patients are more likely to recover faster, while elderly patients (above 80) have instead reduced life expectancies once admitted in the hospital, and usually die in a time window closer to their admission date. Both "tails" of this distribution, despite for different reasons, can be then classified as at low risk of prolonged length of stay, letting us focus our attention on adult and pre/early-retirement individuals.

Resources and Tables

Resource 1. Missing data analysis

```
case_id                            0
Hospital_code                      0
Hospital_type_code                 0
City_Code_Hospital                 0
Hospital_region_code               0
Available Extra Rooms in Hospital  0
Department                         0
Ward_Type                          0
Ward_Facility_Code                 0
Bed Grade                        113
patientid                          0
City_Code_Patient               4532
Type of Admission                  0
Severity of Illness                0
Visitors with Patient              0
Age                                0
Admission_Deposit                  0
Stay                               0
dtype: int64
```

Resource 2. OLS regression with the generated dummy variables as X variables, and the target variable 'Stay_number' as Y - dependent variable

OLS Regression Results

Dep. Variable:	Stay_number	R-squared:	0.073
Model:	OLS	Adj. R-squared:	0.073
Method:	Least Squares	F-statistic:	323.1
Date:	Mon, 10 Mar 2025	Prob (F-statistic):	0.00
Time:	15:28:08	Log-Likelihood:	-6.8348e+05
No. Observations:	318438	AIC:	1.367e+06
Df Residuals:	318359	BIC:	1.368e+06
Df Model:	78		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.069e+11	4.66e+11	-0.658	0.510	-1.22e+12	6.07e+11
Hospital_code_2	-1.987e+10	3.02e+10	-0.658	0.510	-7.9e+10	3.93e+10
Hospital_code_3	5.731e+10	8.71e+10	0.658	0.510	-1.13e+11	2.28e+11
Hospital_code_4	-1.425e+09	2.17e+09	-0.658	0.510	-5.67e+09	2.82e+09
Hospital_code_5	-1.939e+10	2.94e+10	-0.658	0.510	-7.71e+10	3.83e+10
Hospital_code_6	-1.638e+11	2.49e+11	-0.658	0.510	-6.51e+11	3.24e+11
Hospital_code_7	-1.425e+09	2.17e+09	-0.658	0.510	-5.67e+09	2.82e+09
Hospital_code_8	-2.985e+11	4.53e+11	-0.658	0.510	-1.19e+12	5.9e+11
Hospital_code_9	-1.183e+10	1.8e+10	-0.658	0.510	-4.7e+10	2.34e+10
Hospital_code_10	-3.683e+09	5.59e+09	-0.658	0.510	-1.46e+10	7.28e+09

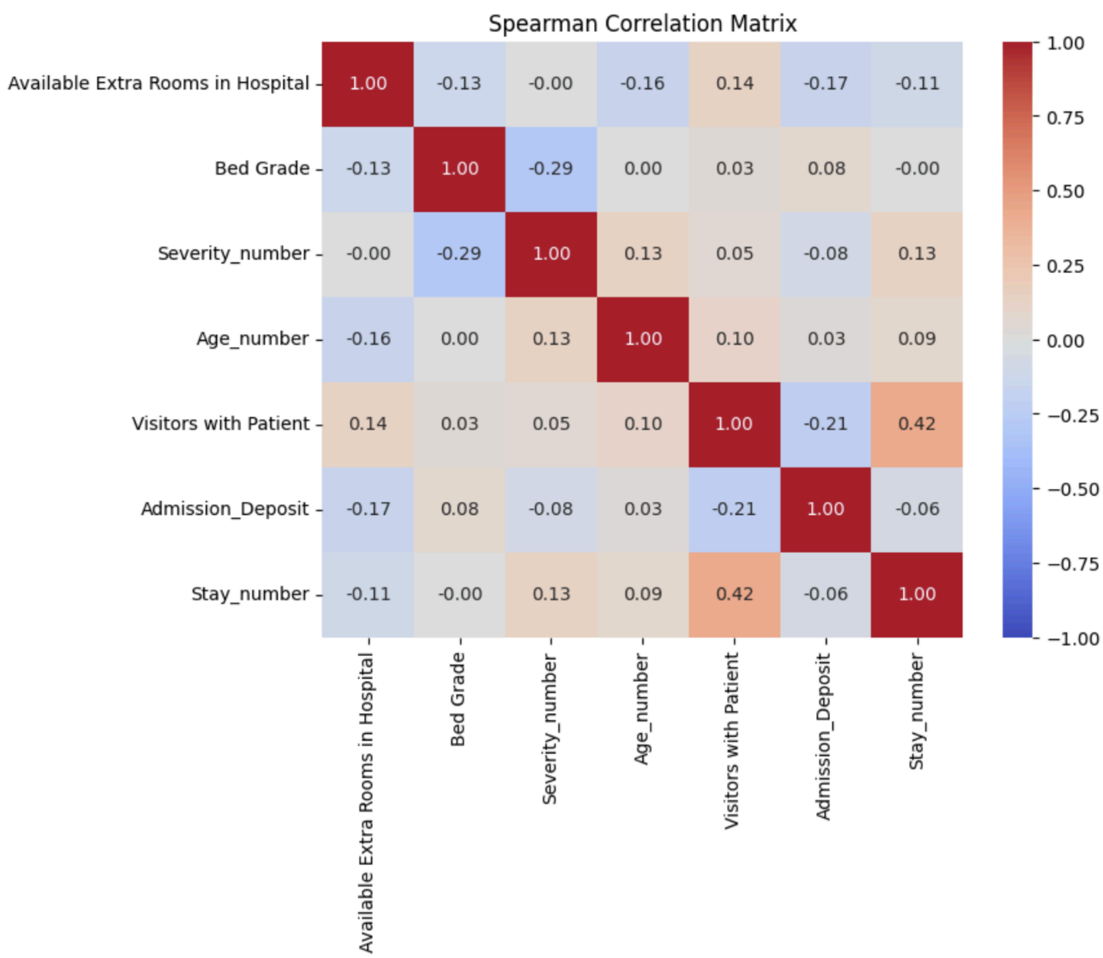
Hospital_code_11	-9.023e+10	1.37e+11	-0.658	0.510	-3.59e+11	1.78e+11
Hospital_code_12	-3.277e+10	4.98e+10	-0.658	0.510	-1.3e+11	6.48e+10
Hospital_code_13	-9.073e+10	1.38e+11	-0.658	0.510	-3.61e+11	1.79e+11
Hospital_code_14	-1.939e+10	2.94e+10	-0.658	0.510	-7.71e+10	3.83e+10
Hospital_code_15	-1.987e+10	3.02e+10	-0.658	0.510	-7.9e+10	3.93e+10
Hospital_code_16	5.731e+10	8.71e+10	0.658	0.510	-1.13e+11	2.28e+11
Hospital_code_17	-3.683e+09	5.59e+09	-0.658	0.510	-1.46e+10	7.28e+09
Hospital_code_18	-3.51e+10	5.33e+10	-0.658	0.510	-1.4e+11	6.94e+10
Hospital_code_19	-1.092e+11	1.66e+11	-0.658	0.510	-4.35e+11	2.16e+11
Hospital_code_20	-9.023e+10	1.37e+11	-0.658	0.510	-3.59e+11	1.78e+11
Hospital_code_21	5.731e+10	8.71e+10	0.658	0.510	-1.13e+11	2.28e+11
Hospital_code_22	-2.061e+10	3.13e+10	-0.658	0.510	-8.2e+10	4.08e+10
Hospital_code_23	-1.638e+11	2.49e+11	-0.658	0.510	-6.51e+11	3.24e+11
Hospital_code_24	-1.939e+10	2.94e+10	-0.658	0.510	-7.71e+10	3.83e+10
Hospital_code_25	-3.683e+09	5.59e+09	-0.658	0.510	-1.46e+10	7.28e+09
Hospital_code_26	-9.023e+10	1.37e+11	-0.658	0.510	-3.59e+11	1.78e+11
Hospital_code_27	-1.092e+11	1.66e+11	-0.658	0.510	-4.35e+11	2.16e+11
Hospital_code_28	-3.446e+10	5.23e+10	-0.658	0.510	-1.37e+11	6.81e+10
Hospital_code_29	-1.425e+09	2.17e+09	-0.658	0.510	-5.67e+09	2.82e+09
Hospital_code_30	5.731e+10	8.71e+10	0.658	0.510	-1.13e+11	2.28e+11
Hospital_code_31	5.731e+10	8.71e+10	0.658	0.510	-1.13e+11	2.28e+11
Hospital_code_32	-1.295e+10	1.97e+10	-0.658	0.510	-5.15e+10	2.56e+10
Hospital_type_code_b	9.79e+09	1.49e+10	0.658	0.510	-1.94e+10	3.89e+10
Hospital_type_code_c	-7.086e+10	1.08e+11	-0.658	0.510	-2.82e+11	1.4e+11
Hospital_type_code_d	-7.89e+10	1.2e+11	-0.658	0.510	-3.14e+11	1.56e+11
Hospital_type_code_e	-1.57e+10	2.39e+10	-0.658	0.510	-6.25e+10	3.1e+10
Hospital_type_code_f	-1.982e+10	3.01e+10	-0.658	0.510	-7.88e+10	3.92e+10
Hospital_type_code_g	-1.216e+10	1.85e+10	-0.658	0.510	-4.84e+10	2.4e+10
City_Code_Hospital_2	-1.394e+11	2.12e+11	-0.658	0.510	-5.54e+11	2.76e+11
City_Code_Hospital_3	2.261e+10	3.43e+10	0.658	0.510	-4.47e+10	8.99e+10
City_Code_Hospital_4	-4.749e+10	7.21e+10	-0.658	0.510	-1.89e+11	9.39e+10
City_Code_Hospital_5	-2.56e+11	3.89e+11	-0.658	0.510	-1.02e+12	5.06e+11
City_Code_Hospital_6	1.149e+11	1.74e+11	0.658	0.510	-2.27e+11	4.57e+11
City_Code_Hospital_7	-1.263e+11	1.92e+11	-0.658	0.510	-5.02e+11	2.5e+11
City_Code_Hospital_9	-1.152e+11	1.75e+11	-0.658	0.510	-4.58e+11	2.28e+11
City_Code_Hospital_10	-6.91e+10	1.05e+11	-0.658	0.510	-2.75e+11	1.37e+11
City_Code_Hospital_11	-2.424e+10	3.68e+10	-0.658	0.510	-9.64e+10	4.79e+10
City_Code_Hospital_13	-3.4e+10	5.16e+10	-0.658	0.510	-1.35e+11	6.72e+10
Hospital_region_code_Y	6.425e+11	9.76e+11	0.658	0.510	-1.27e+12	2.56e+12
Hospital_region_code_Z	2.978e+11	4.52e+11	0.658	0.510	-5.89e+11	1.18e+12
Department_anesthesia	-0.1521	0.024	-6.237	0.000	-0.200	-0.104
Department_gynecology	0.2588	0.022	11.980	0.000	0.216	0.301
Department_radiotherapy	0.3004	0.024	12.262	0.000	0.252	0.348
Department_surgery	0.8529	0.064	13.411	0.000	0.728	0.978
Ward_Type_Q	0.1572	0.031	5.114	0.000	0.097	0.217
Ward_Type_R	0.5791	0.031	18.880	0.000	0.519	0.639
Ward_Type_S	1.3357	0.031	42.872	0.000	1.275	1.397
Ward_Type_T	0.8398	0.064	13.102	0.000	0.714	0.965
Ward_Type_U	-0.4725	0.692	-0.683	0.494	-1.828	0.883

Ward_Facility_Code_B	-1.876e+11	2.85e+11	-0.658	0.510	-7.46e+11	3.71e+11
Ward_Facility_Code_C	-1.001e+11	1.52e+11	-0.658	0.510	-3.98e+11	1.98e+11
Ward_Facility_Code_D	-1.158e+11	1.76e+11	-0.658	0.510	-4.61e+11	2.29e+11
Ward_Facility_Code_E	3.263e+11	4.96e+11	0.658	0.510	-6.45e+11	1.3e+12
Ward_Facility_Code_F	3.558e+11	5.4e+11	0.658	0.510	-7.04e+11	1.42e+12
City_Code_Patient_2.0	-0.1050	0.017	-6.199	0.000	-0.138	-0.072
City_Code_Patient_3.0	0.7763	0.036	21.411	0.000	0.705	0.847
City_Code_Patient_4.0	-0.0663	0.022	-3.023	0.003	-0.109	-0.023
City_Code_Patient_5.0	0.1903	0.020	9.296	0.000	0.150	0.230
City_Code_Patient_6.0	-0.1423	0.031	-4.651	0.000	-0.202	-0.082
City_Code_Patient_7.0	0.1405	0.019	7.229	0.000	0.102	0.179
City_Code_Patient_8.0	0.2529	0.015	16.841	0.000	0.224	0.282
City_Code_Patient_9.0	-0.0094	0.024	-0.395	0.693	-0.056	0.037
City_Code_Patient_10.0	-0.2969	0.027	-11.186	0.000	-0.349	-0.245
City_Code_Patient_11.0	-0.4380	0.082	-5.343	0.000	-0.599	-0.277
City_Code_Patient_12.0	-0.1567	0.031	-5.060	0.000	-0.217	-0.096
City_Code_Patient_13.0	-0.2841	0.053	-5.341	0.000	-0.388	-0.180
City_Code_Patient_14.0	0.2844	0.041	7.005	0.000	0.205	0.364
City_Code_Patient_15.0	0.2805	0.026	10.961	0.000	0.230	0.331
City_Code_Patient_16.0	-0.1332	0.046	-2.920	0.003	-0.223	-0.044
City_Code_Patient_18.0	-0.2520	0.057	-4.391	0.000	-0.364	-0.140
City_Code_Patient_19.0	-0.2830	0.066	-4.290	0.000	-0.412	-0.154
City_Code_Patient_20.0	-0.1810	0.057	-3.171	0.002	-0.293	-0.069
City_Code_Patient_21.0	-0.7555	0.054	-14.034	0.000	-0.861	-0.650
City_Code_Patient_22.0	0.2790	0.104	2.691	0.007	0.076	0.482
City_Code_Patient_23.0	-0.2460	0.037	-6.678	0.000	-0.318	-0.174
City_Code_Patient_24.0	-0.1259	0.110	-1.145	0.252	-0.341	0.090
City_Code_Patient_25.0	-0.0602	0.075	-0.808	0.419	-0.206	0.086
City_Code_Patient_26.0	-0.0345	0.066	-0.522	0.602	-0.164	0.095
City_Code_Patient_27.0	-0.4933	0.076	-6.485	0.000	-0.642	-0.344
City_Code_Patient_28.0	-0.0553	0.092	-0.602	0.547	-0.235	0.125
City_Code_Patient_29.0	-0.2909	0.210	-1.388	0.165	-0.702	0.120
City_Code_Patient_30.0	0.1511	0.180	0.839	0.401	-0.202	0.504
City_Code_Patient_31.0	-0.0922	0.270	-0.342	0.733	-0.621	0.437
City_Code_Patient_32.0	-0.3995	0.287	-1.390	0.165	-0.963	0.164
City_Code_Patient_33.0	-0.3928	0.235	-1.673	0.094	-0.853	0.067
City_Code_Patient_34.0	-0.0541	0.306	-0.177	0.859	-0.653	0.545
City_Code_Patient_35.0	-0.2672	0.518	-0.516	0.606	-1.282	0.747
City_Code_Patient_36.0	-0.2868	0.598	-0.480	0.631	-1.458	0.885
City_Code_Patient_37.0	-0.0797	0.275	-0.290	0.772	-0.618	0.458
City_Code_Patient_38.0	-1.0594	0.845	-1.253	0.210	-2.716	0.597
Type of Admission_Trauma	0.3711	0.009	43.063	0.000	0.354	0.388
Type of Admission_Urgent	-0.0574	0.012	-4.933	0.000	-0.080	-0.035

Omnibus:	70938.357	Durbin-Watson:	1.511
Prob(Omnibus):	0.000	Jarque-Bera (JB):	145489.323
Skew:	1.327	Prob(JB):	0.00
Kurtosis:	4.979	Cond. No.	3.24e+16

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 9.42e-28. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.

Table 1. Heat Map for Correlation Matrix



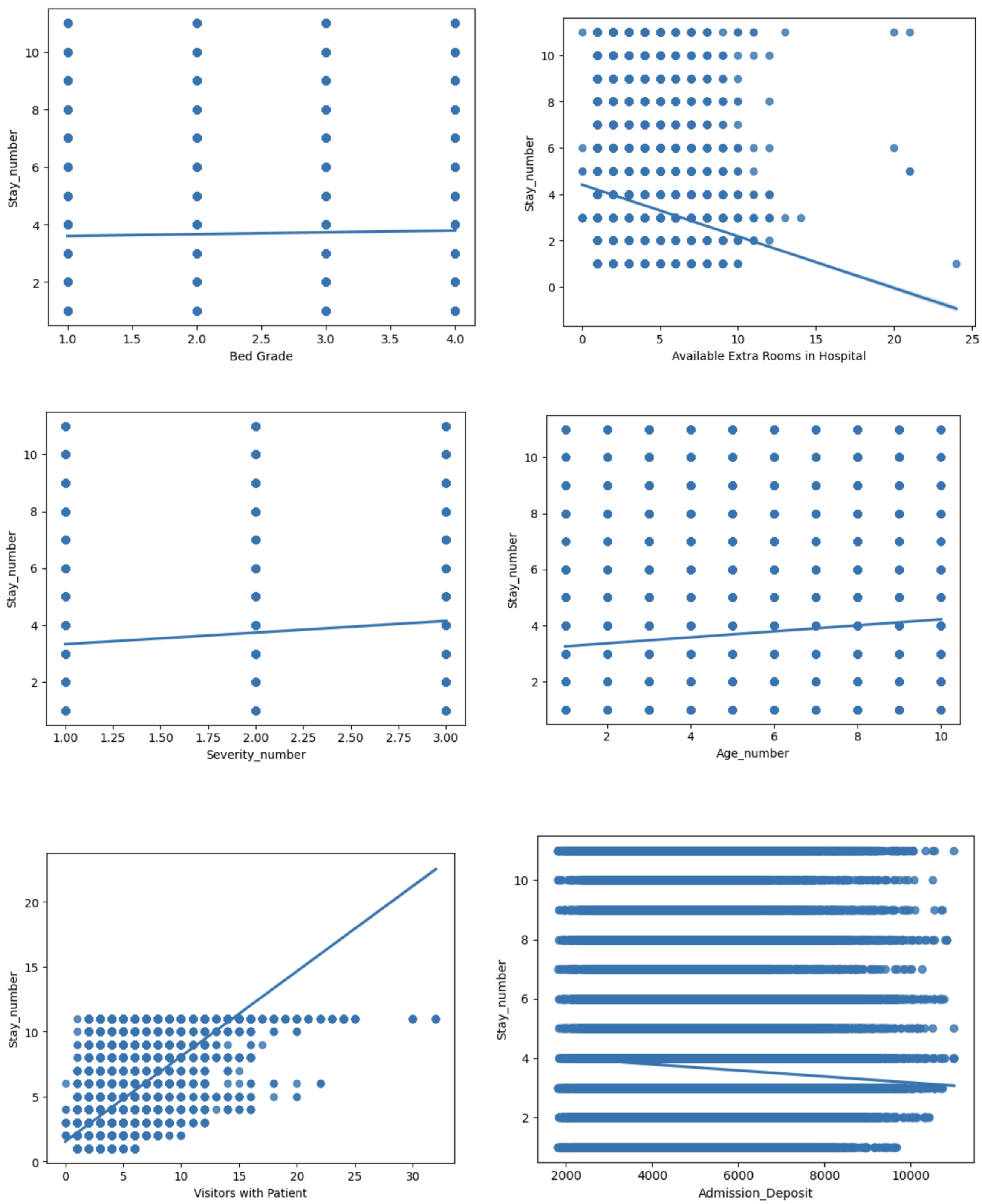


Table 2. Relationship diagram of different ordered variables and LOS

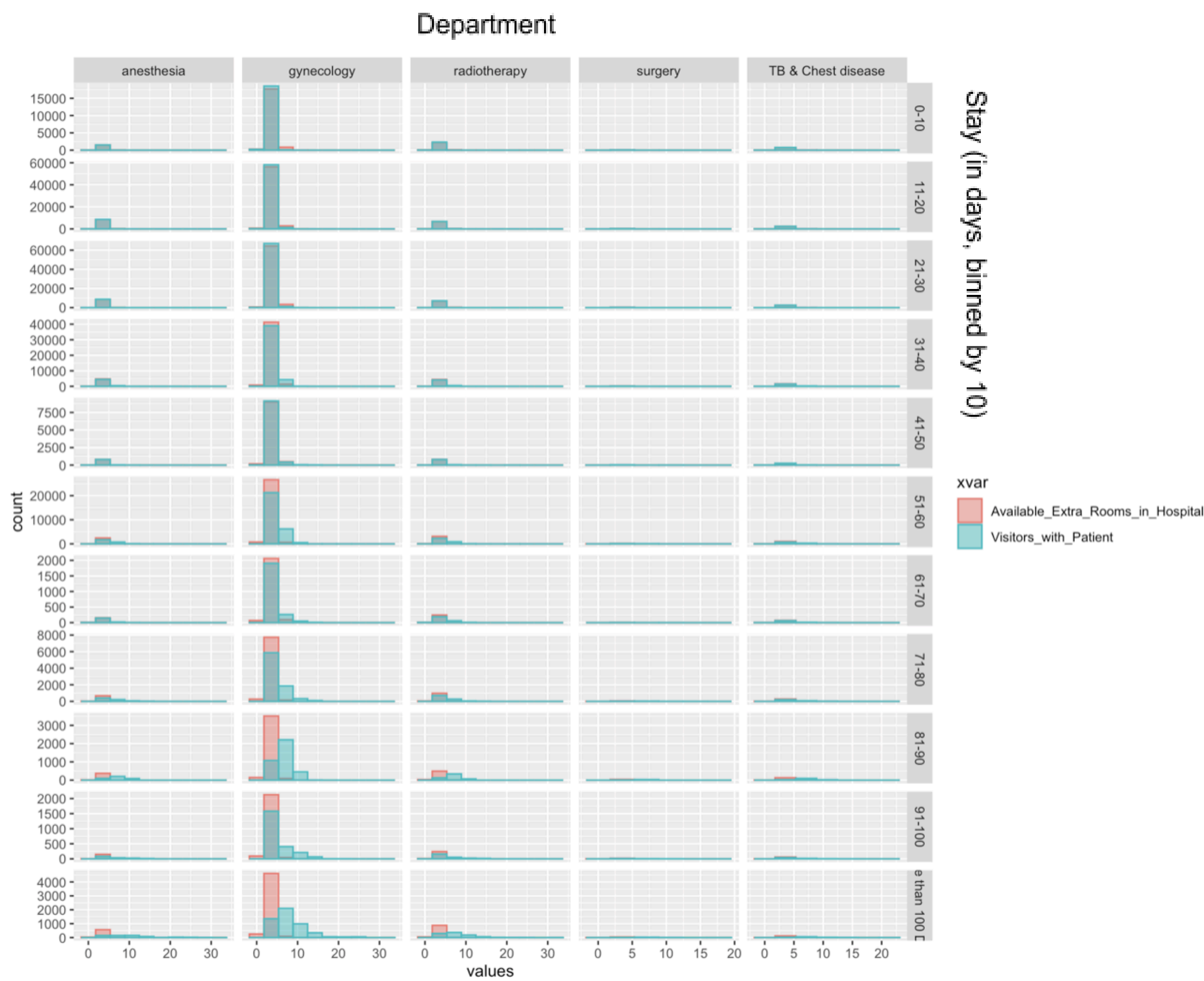


Table 3. Cross distribution table of indicated X-variables by length of stay and department

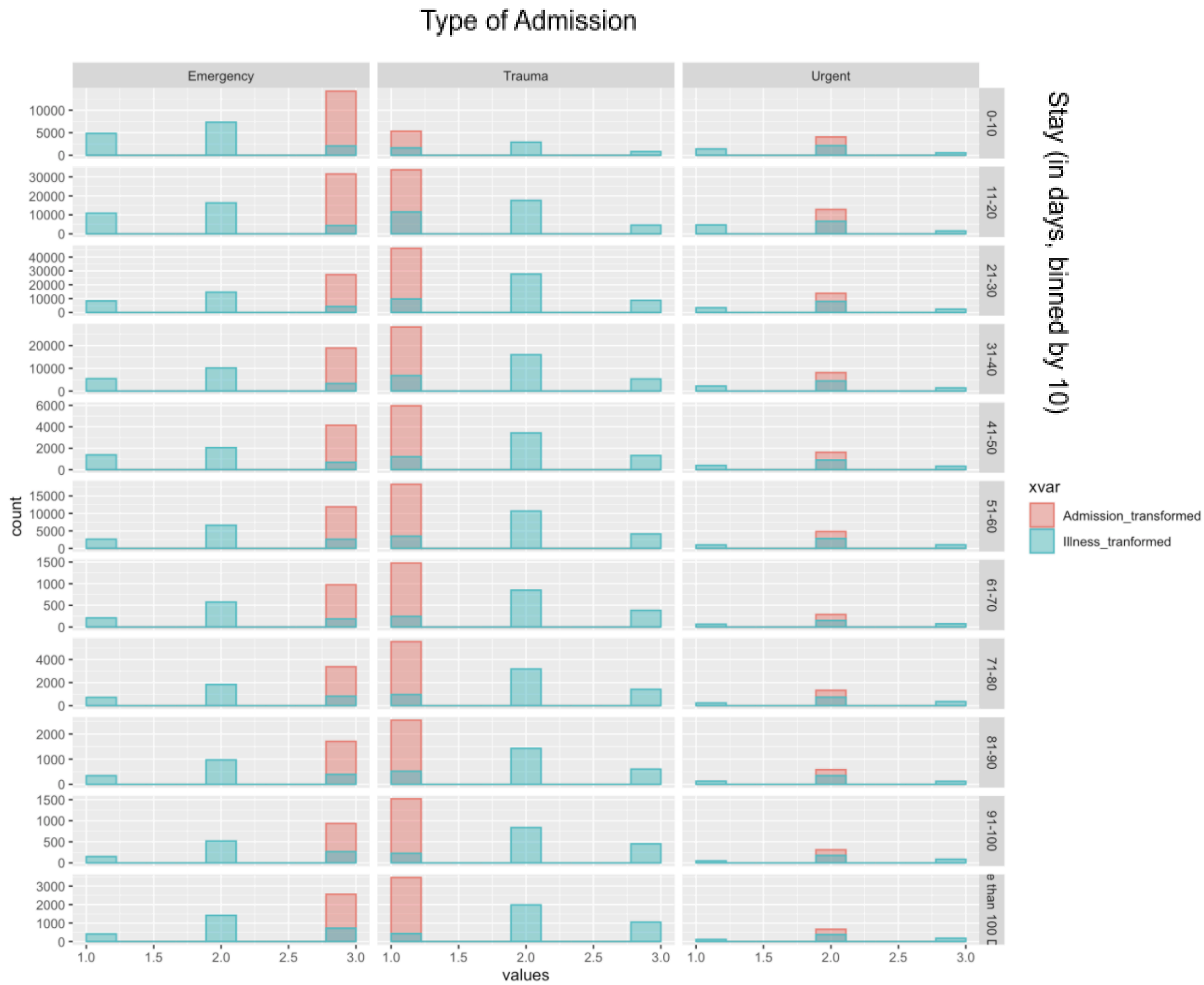


Table 4. Cross distribution table of indicated X-variables by length of stay and type of admission

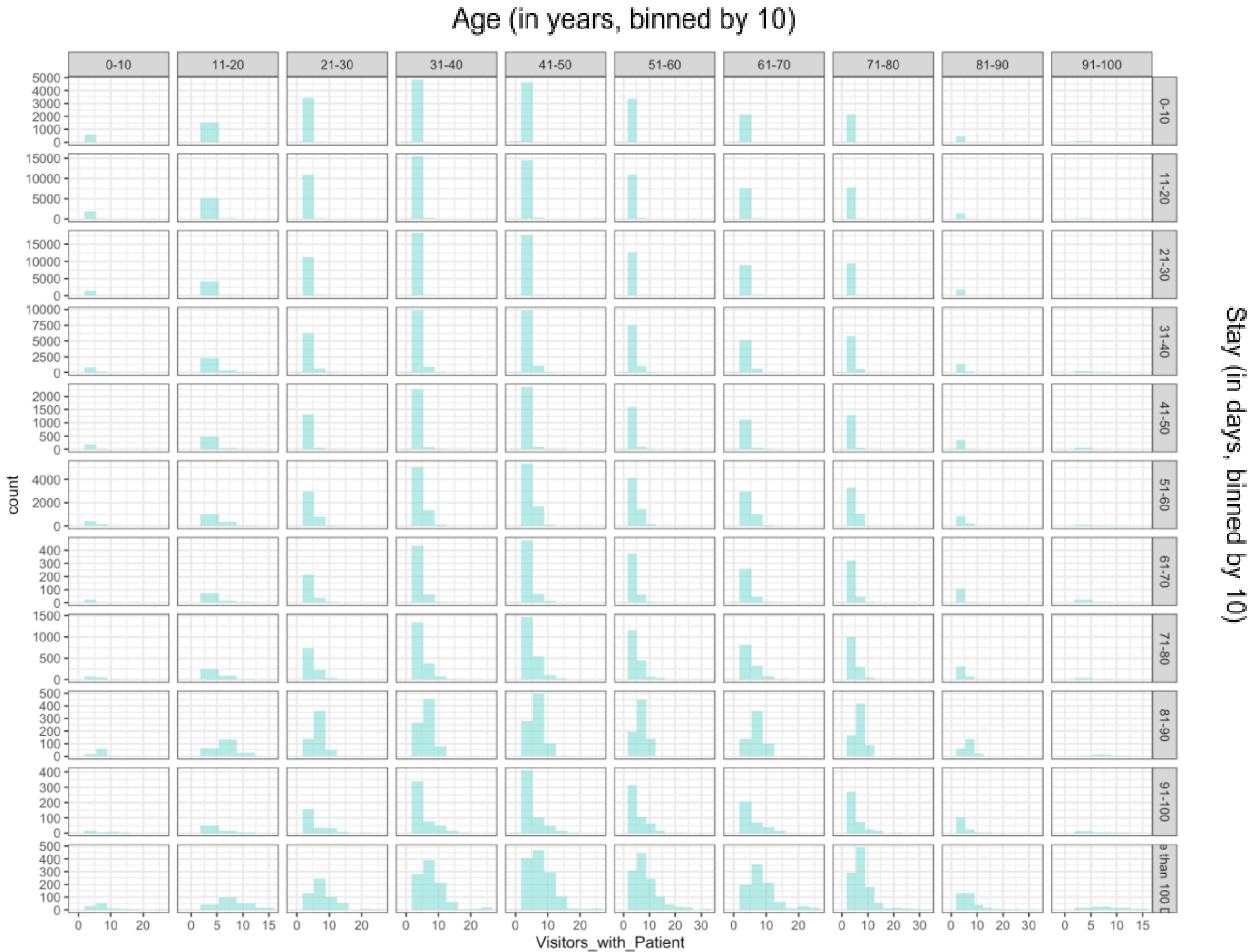


Table 5. Cross distribution table of indicated visitors by length of stay and age