HW2 : *Elective Surgery Schedule*

Ayush Agarwal

Q1: Provide some reasons as to why predictive modeling can be helpful here. We may need to justify the use of resources on an analytics-oriented solution, so please think broadly about ways that prediction can be helpful.

A1 : Predictive modelling becomes an essential tool for healthcare institutions such as Vanderbilt University Medical Centre when effective resource management and optimal patient care are vital. The following aspects from the case study demonstrate how this strategy improves patient outcomes in addition to streamlining operations:

1. **Resource Optimisation**: The case study highlights problems with the effectiveness of staffing. By predicting patient numbers, predictive analytics can help with staff scheduling to avoid overstaffing and maximise the use of available resources.
2. **Taking Care of Emergencies**: The case study's unpredictable emergency scenarios emphasize the necessity of flexible scheduling. By anticipating and allocating room for these circumstances, predictive models can assist make sure they are handled without causing significant disruptions.
3. **Staff Leave Management**: The case study highlights the difficulties in staff leave administration. Staff satisfaction can be increased without sacrificing patient care by using predictive analytics to strike a balance between staff availability and their need for paid holidays and research time.
4. **Effective Equipment Management**: Keeping an eye on the supply of sterile equipment can be difficult, as the case study illustrates. By using predictive modelling to forecast surgery volumes, equipment supply and demand can be more effectively managed.
5. **Enhanced Operational Efficiency**: As a whole, the case study shows that operational efficiency needs to be raised. Predictive analytics may streamline several elements of hospital operations, from staffing to equipment management, resulting to cost savings and better patient care.

Q2: Using the data, explain:

1. **Whether and how the number of scheduled cases is predictive of actual final case volume**
2. **Whether the predictive power of scheduled cases changes as the surgery date nears (e.g., whether T-28 is differently predictive than, say, T-5)**
3. **How, if at all, total surgical volume differs by day of week**

A2:

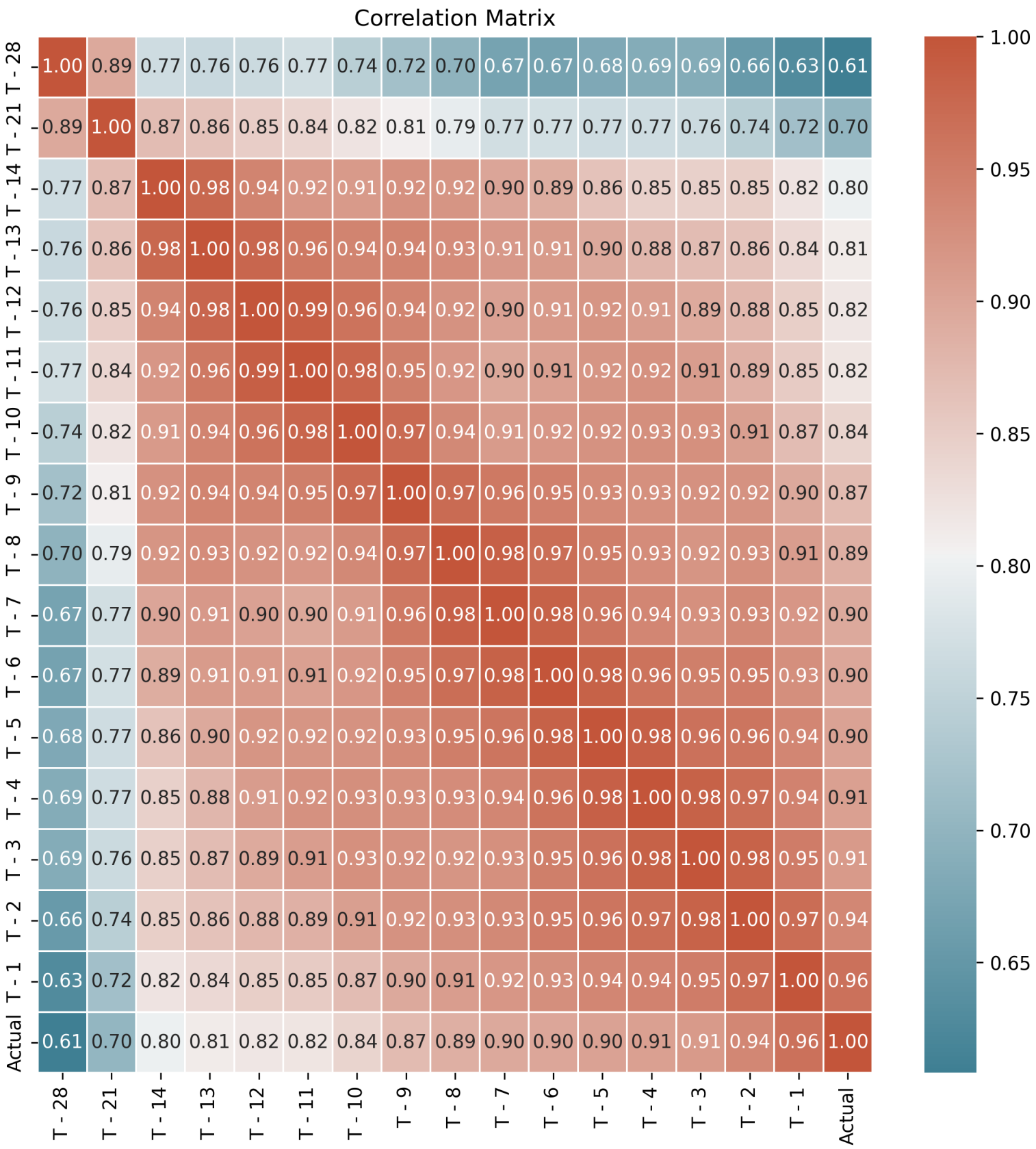
1. The number of scheduled cases is predictive of actual final case volume. The number of scheduled cases given T-28, T-27.. T-x, and so on gives the number of cases scheduled x days prior to the surgery day. We can write the actual final case volume using an equation given below

$$y = cases(x) + e$$

   where y is the actual final case volume and cases(x) represent scheduled cases on x day prior and e is the emergency cases with respect to xth day. Intuitively and by visual data analysis as well, it can be clearly assumed that T-1 holds the highest predictive power. I have verified this using correlation matrix between predictor variables and actual final case volume. T-1 has highest correlation with y , followed by T-2, T-3…..T-28

2. Yes the predictive power of scheduled cases changes as the surgery date nears. We can analyze this using two approaches
   - **Correlation Matrix Approach** : Below is the correlation plot among actual value and predictor variables (T-x). From the plot , we can clearly see, T-1 has highest correlation with actual followed by T-2, T-3, ….T-28 and hence we can conclude T-1

has highest predictive power than T-5 or T-28

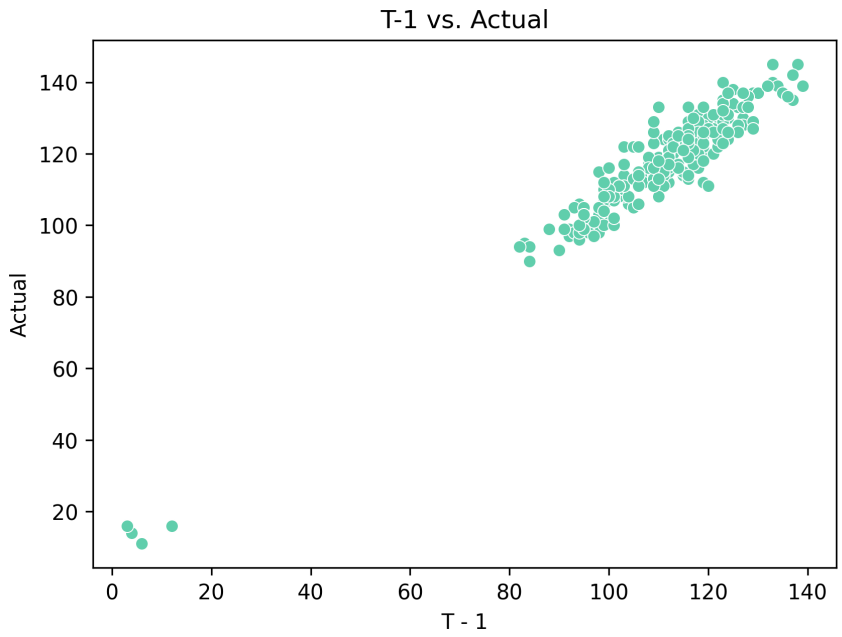


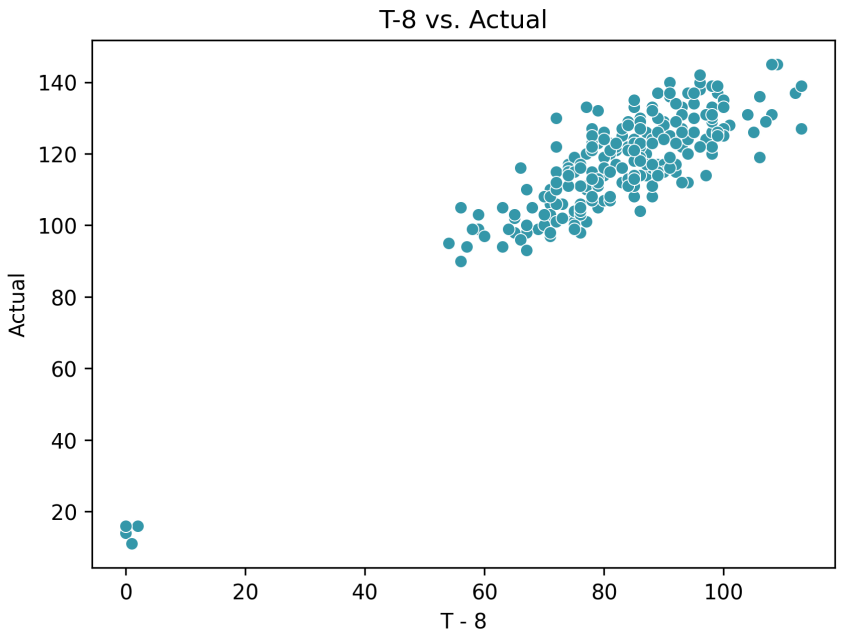
- Regression Analysis : Fit a linear regression model against Actual vs T-x predictors separately and analysed their adjusted R square value. Below is a comparison of T-1, T-5, T-28

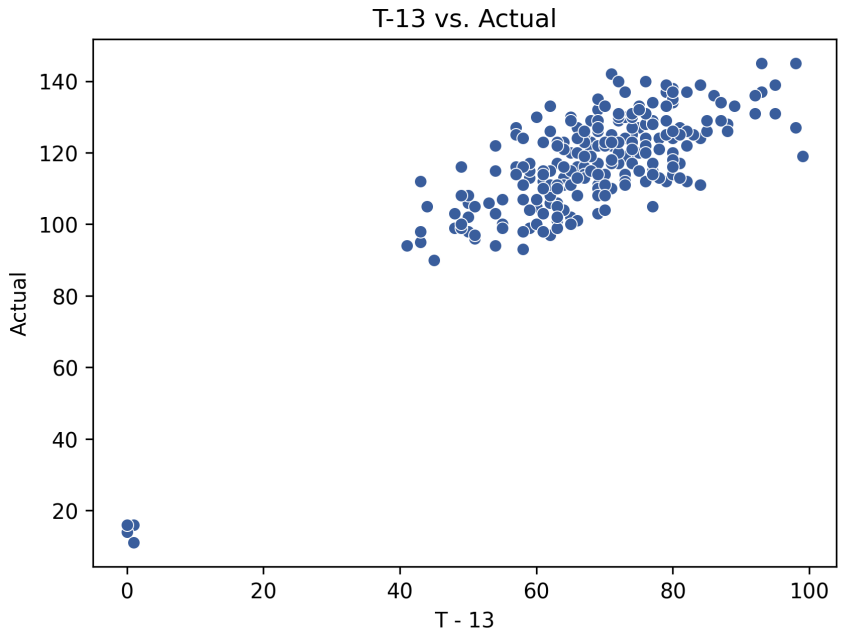
Predictor	Adjusted R2
T-1	0.93
T-5	0.81
T-28	0.37

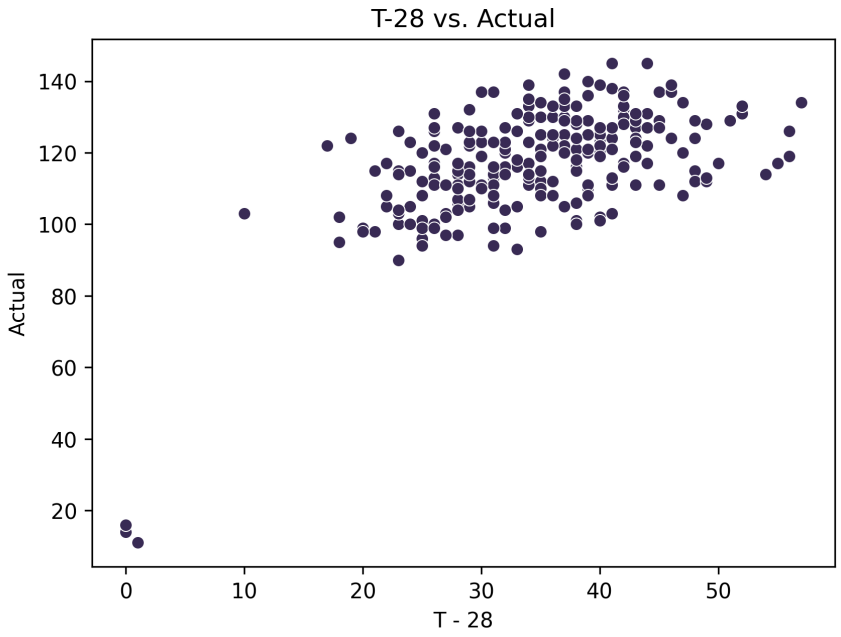
It clearly depicts that as x decreases, predictive power of the model increases.

Below scatterplots also demonstrate predictive power of T-x wrt Actual









3. At Vanderbilt University Medical Centre, the overall surgery volume varies depending on the day of the week. Thursdays are the busiest day for surgery, with the greatest mean number of cases (124). On the other hand, Fridays had a mean of 106 fewer surgical cases than other days, indicating that fewer surgeries are scheduled for the weekend. There is a moderate consistency in the surgery volumes on the other days, which are Monday through Wednesday, with averages that are quite similar, ranging from 116 to 119 cases. There could be other reasons for this variance in the number of surgeries performed, such as patient convenience, staff availability, and scheduling preferences.

Day of Week	Mean number of Cases
Monday	116
Tuesday	119
Wednesday	117
Thursday	124
Friday	106

Q3: **Show us the output of a predictive model that you think will be most helpful for us to accomplish any of the goals you mentioned in point 1. Also briefly describe what kind of model you are using and any important elements of the model that you think we need to know.**

A3: In my opinion, a linear regression model that focuses on predictors from T-28 to T-7 days before surgery is the best fit for the operational requirements of Vanderbilt University Medical Centre. This model is good at predicting important operational parameters; it can accurately capture around 80% of the variance in real surgeries. The ability to provide precise forecasts up to one week beforehand facilitates the effective distribution of resources, encompassing personnel and equipment administration. Managing staff leave, responding to emergencies, and guaranteeing the availability of essential equipment all depend on this forethought. The model's usefulness in predicting surgical demands is highlighted by the notable influence of the T-7 predictor in particular.

Key Findings:

- The model is able to capture ~80 % of the variance in actual surgeries performed using the given T-x predictor variables
- The most recent predictor variable (T-7) has the highest impact on the surgeries performed

OLS Regression Results

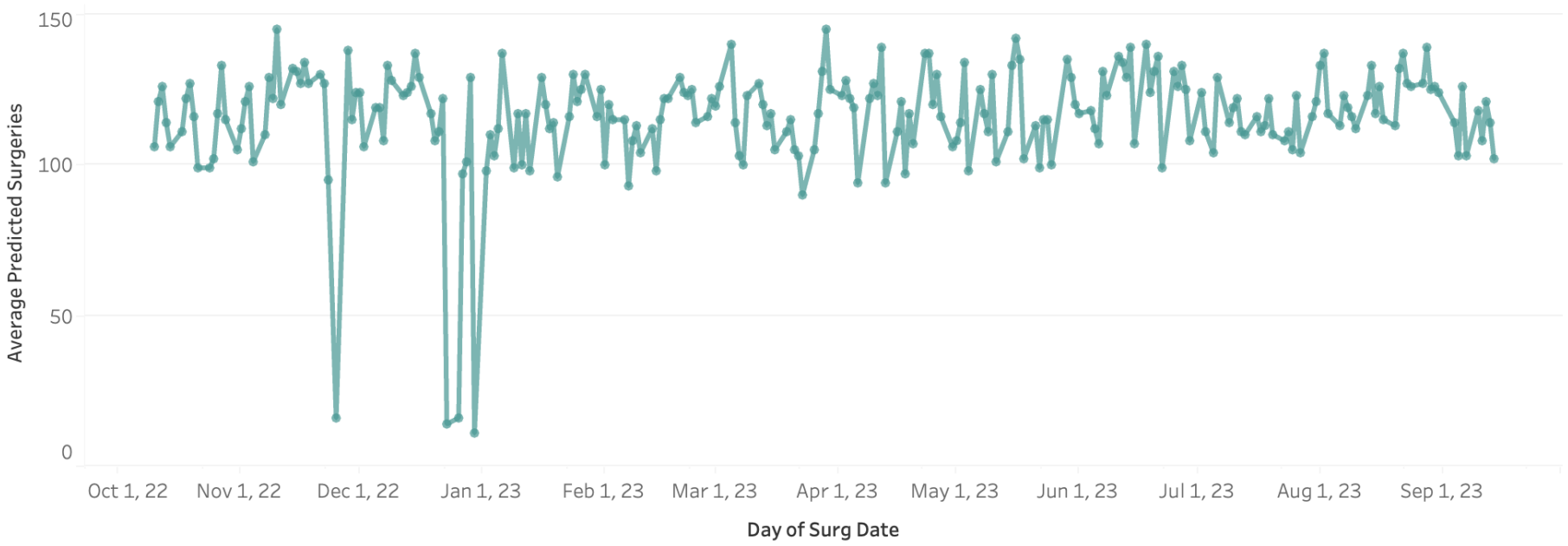
Dep. Variable:	Actual	R-squared:	0.814
Model:	OLS	Adj. R-squared:	0.806
Method:	Least Squares	F-statistic:	100.7
Date:	Sat, 13 Jan 2024	Prob (F-statistic):	3.21e-78
Time:	21:03:42	Log-Likelihood:	-830.26
No. Observations:	241	AIC:	1683.
Df Residuals:	230	BIC:	1721.
Df Model:	10		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	31.0812	2.764	11.247	0.000	25.636	36.526
T - 28	-0.0940	0.123	-0.761	0.447	-0.337	0.149
T - 21	0.1628	0.131	1.246	0.214	-0.095	0.420
T - 14	0.0369	0.191	0.193	0.847	-0.340	0.414
T - 13	-0.5722	0.281	-2.036	0.043	-1.126	-0.019
T - 12	0.3446	0.321	1.072	0.285	-0.289	0.978
T - 11	-0.1966	0.314	-0.626	0.532	-0.815	0.422
T - 10	0.2834	0.248	1.142	0.254	-0.205	0.772
T - 9	0.1275	0.231	0.552	0.581	-0.328	0.583
T - 8	0.0375	0.237	0.158	0.874	-0.429	0.504
T - 7	0.8477	0.185	4.579	0.000	0.483	1.213

Q4: Please mock up a sample dashboard / graphic that we could update and use on a daily basis that you think would help us. Consider what information you think the team would want to see regarding upcoming predicted volume. Don't worry if you have the right data available for this example – you can use what you have or just make up data points to fill in the dashboard / graphic.

A4: I have created a dashboard that will help VUMC to track the predicted surgeries of the upcoming week as well as monitor the error rate of the model based on historical and current captured data. The first plot displays the average surgeries predicted by a model on a particular day. This will help the VUMC to prepare in advance for the upcoming workload. The second plot shows the prediction error of the model. Using this the data science team of VUMC can keep track of the model's performance. They can consider updating or training the model again when error rates are constantly high above some set threshold for a few days or weeks.

Daily Predicted Surgeries



Prediction Error