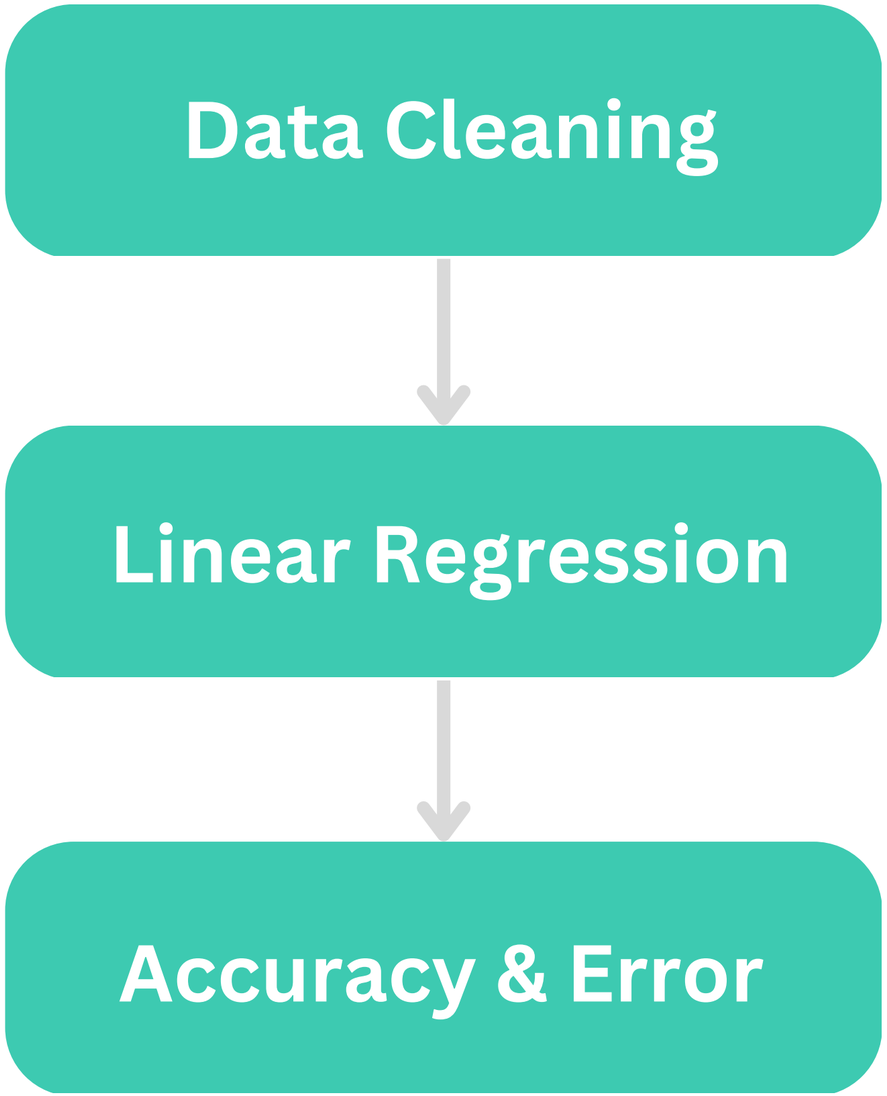
Machine Learning COVID-19 Hospital Capacity

Grace Avalos

Process



Cleaning the Data

- Replace values of -999,999 with 3
- Remove rows where inpatient_beds_used_covid_7_day_sum = "NA"
- Remove unnecessary columns

```
covid_data <- COVID.19_Reported_Patient_Impact_and_Hospital_Capacity_by_Facility

sample<- data.frame(covid_data$inpatient_beds_used_covid_7_day_sum, covid_data$hospital_pk, covid_data$city,covid_data$state, covid_data$is_metro_micro)

sample$covid_data$inpatient_beds_used_covid_7_day_sum[sample$covid_data.inpatient_beds_used_covid_7_day_sum == -999999] <- 3

subset(sample,covid_data$inpatient_beds_used_covid_7_day_sum is.NA)
```

Linear Regression

Predict **inpatient_beds_used_covid_7_day_sum** values for the next 12 weeks for:

- A specific hospital key selected ("hospital_pk")
- A specific city selected ("city")
- A specific hospital type selected ("is_metro_micro")
- A specific state selected

```
# create training set
set.seed(88)
split = sample.split(covid_data$inpatient_beds_used_covid_7_day_sum, SplitRatio = 0.80)
split
SampleTrain = subset(sample, split==TRUE)
SampleTest = subset(sample, split==FALSE)

# multiple variable linear regression
lm(covid_data.inpatient_beds_used_covid_7_day_sum ~ covid_data.hospital_pk + covid_data.city +
     covid_data.is_metro_micro + covid_data.state, data=sample)
```

Linear Regression

Accuracy:

```
SampleReg = lm(covid_data.inpatient_beds_used_covid_7_day_sum ~ covid_data.hospital_PK + covid_data.city +
                 covid_data.is_metro_micro + covid_data.state, data=sample)
summary(SampleReg)
```

Margin of error:

```
# margin of error
SSE = sum((SampleTest$covid_data.inpatient_beds_used_covid_7_day_sum - lm)^2)
SST = sum((SampleTest$covid_data.inpatient_beds_used_covid_7_day_sum - mean(Sample$covid_data.inpatient_beds_used_covid_7_day_sum))^2)
1 - SSE/SST
```

Linear Regression Explanation

Relationship between **inpatient_beds_used_covid_7_day_sum** and other variables

Can find best fit line linear line between **inpatient_beds_used_covid_7_day_sum** and independent variables

Supervised machine learning model

Strength of **inpatient_beds_used_covid_7_day_sum** and other variables