Alec Engel
April 25, 2022

Final Write-Up for New Children’s Hospital Location in India

Project Scope

India and their children are in dire need of a new Children’s Hospital to care of the ill. There has been growing concern across India for lack of children’s health care facilities and the poor environmental safety (air quality and noise decibels) where existing facilities are located. At young ages, it is important to protect children’s sensitive lungs by allowing them to breath the highest quality air and protect their sensitive ear drums by not exposing them to extremely high decibel levels. These factors are even more important when dealing with sick children in need of hospital care. With this in mind, India’s government has set aside money based on their taxation budget for a new state of the art facility. India needs to focus on locating their new Children’s Hospital in an area with clean air and low decibel levels.

BigAlData Consulting Company has been tasked by India’s healthcare system to find the most viable location for the new Children’s Hospital with environmental health and safety of the children as the number one priority. Location, location, location is the key focus for this project. After five weeks of research and 100 hours, BigAlData is presenting this project write-up to recommend the preferred location of building this new Children’s Hospital.

Project proposal was submitted by March, 21 2022, project methods were submitted by March 28, 2022, project data cleaning and preparation was performed prior to April 11, 2022, analytical model building took place prior to April 25, 2022, and this final project write-up with decision analysis is being presented prior to April 28, 2022. BigAlData used air quality index (AQI) and noise decibel levels gathered from stations across India between 2015 and 2020 to produce statistical research methods and visualizations to help support the hospital location decision. This collection of data will help predict future clean air and noise levels to recommend the selection of one city. Valuable descriptive sets of AQI and decibel level data also helped BigAlData Consulting Company perform descriptive, predictive, and prescriptive analytical techniques to produce a recommendation of the best location of the future new Children’s Hospital. With that being said, BigAlData is being tasked with finding the purest city to build the new hospital, not to find the precise lot that may be for sale within the resulting city to build on.

Deliverable Techniques

BigAlData Consulting Company created a dashboard filled with visualizations to help support the recommended location of where to build the new Children’s Hospital with environmental health aspects as top priority. These visualizations combined raw data for individual stations across India in order to target cities with best air quality and lowest decibel levels. 2015 to 2020 trend charts were also created with given data to show progress or lack thereof for specific cities so that a prediction can be made for future environment conditions once the hospital is built and children are able to become patients.

R programming was used for model building of predictive analytical modeling techniques to prove strength in the data and reliability of outcomes. The 2015 to 2020 data trends obtained will be used to help predict conditions in the future. The predictive data gathered from R programming will also further strengthen the end decision based on the proven strength of predictability with data shown in each model.

Excel was also used to create pivot tables and charts to help with descriptive as well as prescriptive statistics of each set of data. By formulating a city's mean, mode, and standard deviation for AQI and decibel levels, BigAlData was able to produce reliable expected probabilities for future health ratings. These probabilities further allowed comparisons of top and bottom potential fits while exemplifying additional strength in the given recommendation.

Data Cleansing

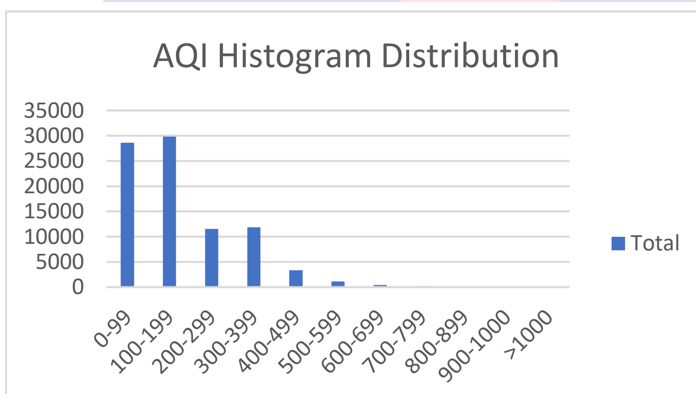
BigAlData performed cleanup, consolidation, and normalization on the eight data sets provided for India's Air Quality Index and decibel levels. VLOOKUP's, date modifications, concatenations, and replacing missing data were all part of the cleanup process to consolidate the AQI and decibel datasets into one a piece. For each dataset, N/As were then removed to perform data analysis since a missing AQI numbers lead to insufficient variable calculations and there were only three missing decibel numbers. The focus on this cleansing exercise was geared towards monthly findings per city in which a collection station was located. Descriptive analytics was then performed on the remaining two datasets to get a view into what has happened in the past with India's AQI and decibel levels. Supporting data analysis packages and visualizations within excel were ran to spot variances and backup our findings.

Using descriptive statistics and data analysis packages within excel, a summary chart was created with all AQI data numbers. Mean AQI was so much higher than median and mode which signified the possibility for some fairly significant outliers within the dataset. The city of Ahmedabad shows to be extremely volatile with 24 severe AQI readings including the two maximum AQI readings of 2,049 and 1,917 in February of 2018. We will be sure to steer clear of Ahmedabad moving forward when looking for hospital locations as we can't risk another month like this with such poor AQI for the children.

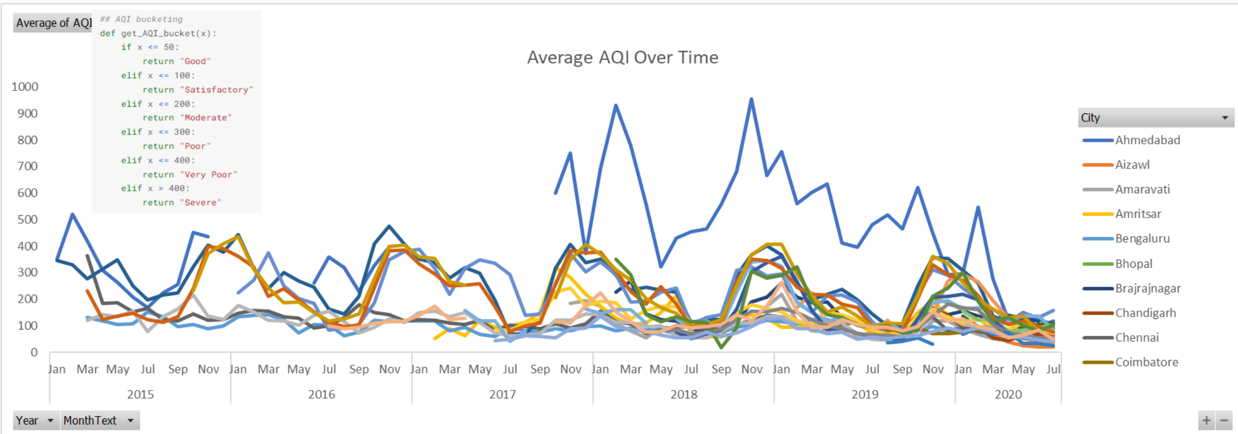
AQI	
Mean	179.7493
Standard Error	0.445167
Median	132
Mode	104
Standard Deviation	131.3243
Sample Variance	17246.08
Kurtosis	8.532544
Skewness	1.930088
Range	2041
Minimum	8
Maximum	2049
Sum	15642682
Count	87025
Largest(2)	1917
Smallest(2)	10

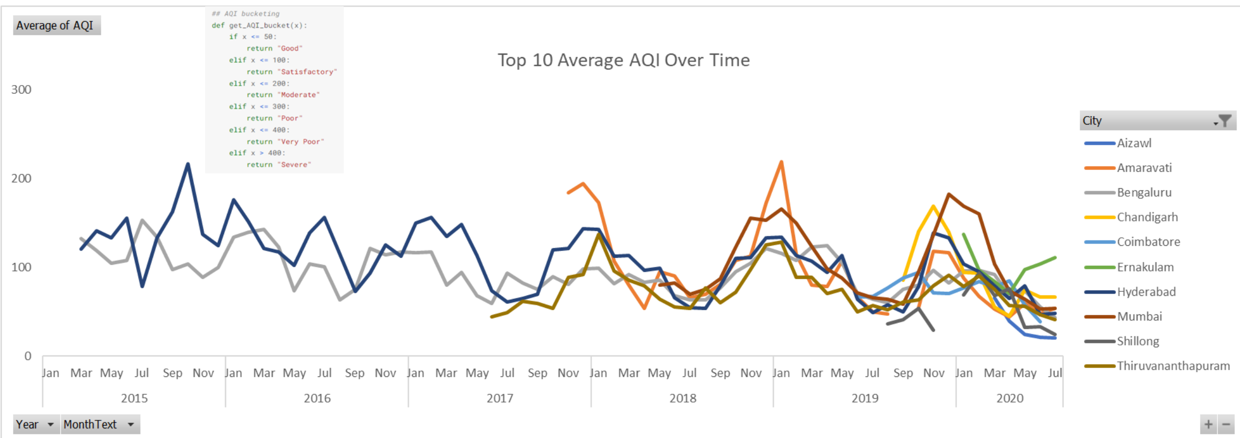
AQI rankings and their standard deviations of average AQI levels over the five years of data collected were then used within an excel pivot table to generate a visualization of top and bottom overall performing cities. The top four cities of Aizawl, Shillong, Coimbatore, and Thiruvananthapuram seem to have low deviations as well as often returning satisfactory or better AQI ratings consistently. Again, with this chart we can see how far behind Ahmedabad in particular as well as Delhi and Lucknow are in terms of AQI health. A histogram was also created in excel to show the positively skewed distribution which again supports the findings of several extreme outliers, primarily with Ahmedabad.

Cities	Total AQI Ranking	Total Average of AQI	Total StdDev of AQI
Aizawl	1	34.8	20.0
Shillong	2	53.8	28.4
Coimbatore	3	73.0	18.1
Thiruvananthapuram	4	74.8	28.0
Bengaluru	5	91.7	49.6
Ernakulam	6	92.4	19.4
Amaravati	7	95.3	56.1
Chandigarh	8	96.5	57.2
Hyderabad	9	100.4	50.1
Mumbai	10	103.1	59.9
Kochi	11	104.3	34.0
Chennai	12	113.6	68.6
Visakhapatnam	13	117.3	54.9
Amritsar	14	119.9	77.0
Kolkata	15	132.7	98.8
Bhopal	16	132.8	56.0
Jaipur	17	134.8	63.2
Guwahati	18	140.1	112.8
Brajrajnagar	19	150.3	70.4
Jorapokhar	20	159.3	84.7
Talcher	21	172.9	104.9
Gurugram	22	211.0	116.9
Patna	23	215.6	120.9
Lucknow	24	216.5	124.8
Delhi	25	238.7	130.3
Ahmedabad	26	452.1	311.7
Grand Total		**179.7**	**131.3**



A city's average AQI over time was then shown in a line chart. Although Aizawl, Shillong, Coimbatore, and Thiruvananthapuram appear to be at the low end of AQI average measures, only Thiruvananthapuram consisted of records dating past 2019 (having readings since 2017). This chart generated shows potential patterns in our AQI dataset. An additional glaring pattern jumps out, aside from our predetermined outlier city of Ahmedabad. AQI seems to have increases from the six months of October through March and significant decreases from the six months of April through September. Additional exploratory analysis will need to be performed on these findings to take extra precautions for our patients during these high AQI time periods regardless of where the Children's Hospital is built.



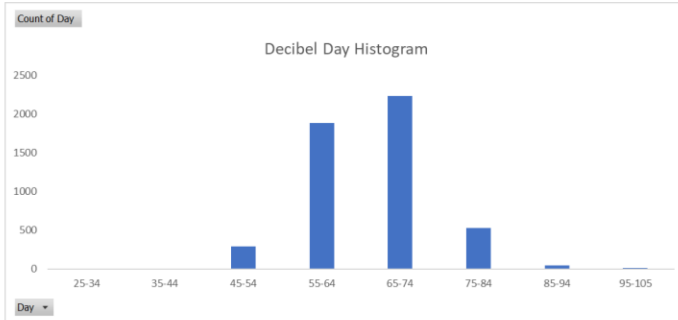


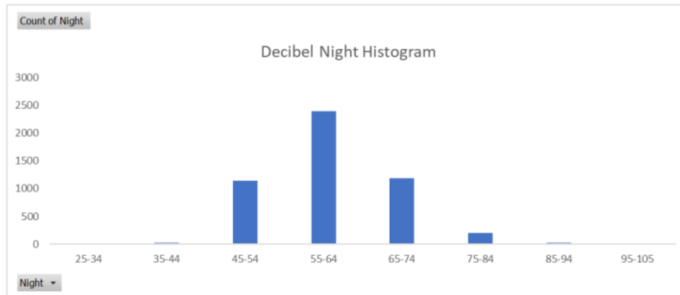
Using descriptive statistics, a summary chart was created with all decibel day and night data. Decibel levels during the day seemed to be higher than at night which makes sense with more human activity. Both time periods were still tracked since there may be children staying overnight in the new hospital. Curiosity was also peaked here to see if there was a correlation between day and night noise in the cities and if so, did they have a significant relationship? After performing regression statistics on the decibel dataset in excel, a strong r value of .88 (close to 1) and a low p-value less than .05 shows a strong relationship. These outcomes proved that a strong correlation and significant relationship existed with decibel levels in a city day and night. Decibel levels when viewed through a histogram also appear normally distributed whether recorded during the day or at night. This again

Day		*Night*	
Mean	65.83063	Mean	60.43083
Standard Error	0.104812	Standard Error	0.115638
Median	66	Median	59
Mode	64	Mode	57
Standard Deviation	7.407644	Standard Deviation	8.17273
Sample Variance	54.87319	Sample Variance	66.79352
Kurtosis	0.795961	Kurtosis	1.396504
Skewness	0.279136	Skewness	0.723751
Range	69	Range	78
Minimum	31	Minimum	25
Maximum	100	Maximum	103
Sum	328824	Sum	301852
Count	4995	Count	4995
Largest(2)	100	Largest(2)	100
Smallest(2)	32	Smallest(2)	29
SUMMARY OUTPUT			
Regression Statistics		P-value	
Multiple R	**0.883252**	2.06677E-14	

reiterates there being a correlation among the times of day without many outliers.

Based on an additional heatmap of yearly decibel level averages, we can see again the correlation between day time and night time decibel levels as their rankings are the same. Delhi, Bengaluru and Lucknow seem to have lowest overall decibel levels while Chennai, Navi Mumbai and Kolkata have the highest. One key takeaway from this chart was the increase of decibel levels for all cities during the daytime seemingly from 2015 onwards.





	2011	2012	2013	2014	2015	2016	2017	2018	Total Average of Day
Cities	Average of Day	Average of Day	Average of Day	Average of Day	Average of Day	Average of Day	Average of Day	Average of Day	
Delhi	59.8	58.5	59.2	59.9	65.6	63.2	62.0	61.3	61.8
Bengaluru	60.5	59.4	60.4	60.8	63.6	63.6	64.5	64.9	62.9
Lucknow	61.6	62.0	63.3	65.0	69.1	69.4	66.2	64.7	65.9
Hyderabad	65.1	64.1	64.4	66.0	67.6	68.1	67.9	67.1	66.8
Mumbai	66.3	65.6	66.2	67.8	67.7	67.5	66.9	63.5	66.4
Kolkata	61.7	61.9	63.5	64.2	68.2	70.2	67.9	66.6	66.5
Navi Mumbai	68.0	68.7	68.6	68.8	68.3	70.3	70.1	70.9	69.2
Chennai	68.6	69.4	69.8	69.8	71.1	72.7	70.6	67.8	70.2
Grand Total	**63.4**	**63.1**	**63.9**	**64.8**	**67.6**	**67.8**	**66.6**	**65.2**	**65.8**

									Total Average of Night
Cities	Average of Night	Average of Night	Average of Night	Average of Night	Average of Night	Average of Night	Average of Night	Average of Night	
Delhi	55.9	54.7	55.4	55.3	60.9	57.6	57.0	57.0	57.2
Bengaluru	54.4	52.2	53.7	54.6	58.0	58.5	59.9	60.5	57.4
Lucknow	54.4	54.0	55.0	56.5	62.2	62.9	60.0	58.6	59.0
Hyderabad	58.3	57.3	57.6	59.2	61.7	63.0	63.0	62.3	61.1
Mumbai	61.0	61.1	61.7	62.9	62.8	63.6	63.4	59.7	62.2
Kolkata	55.6	56.1	59.6	59.5	64.1	66.8	64.0	63.5	62.4
Navi Mumbai	59.4	58.1	56.9	59.0	64.2	68.1	69.9	64.4	62.5
Chennai	59.5	59.8	60.4	61.4	65.0	67.4	65.9	64.0	63.8
Grand Total	**57.0**	**56.4**	**57.5**	**58.4**	**62.1**	**62.9**	**61.9**	**60.9**	**60.4**

Data Sampling & Validation

BigAlData next performed additional cleanup and sampling on the remaining data sets for India's Air Quality Index and decibel levels. For realistic results; regression statistics, correlation, covariance, confidence intervals, and sample size determinations were all calculated. Supporting data analysis packages, pivot tables and formulas within excel were then used to produce these findings.

Using descriptive sampling statistics, a summary chart and ideal sample size was generated with AQI data numbers. The goal here was to produce a sample standard deviation and mean that were as close as possible to the population standard deviation mean (Central Limit Theorem). A 95% confidence interval was used for production accuracy. With a sample mean of 178.78 and sample error of 1.63 amongst a sample size of 25,000, BigAlData was able to come up with a sample mean AQI range between 177.15 and 180.41 for all cities. With the data analysis sampling tool, it was also determined to achieve a maximum error of 2 AQI, a sample size of just 16,563 data records would be required. With 67.29% of cities scoring satisfactory, moderate or good AQI, a sample size of 2,114 would be needed to predict with 95% accuracy that a rating would result in one those top three categories. This all proves how reliable this strong dataset is moving forward with predictive and prescriptive analytic strategies.

Sample Size Determination --> Confidence Interval for Mean		
alpha (α)	0.05	Enter values in red
Standard Deviation	131.3243	Sample size calculated below
Desired Maximum CI Width	2	
z	1.96	
Sample Size	16563	<-Rounds up to nearest integer

Sample Size Determination --> Confidence Interval for Proportion		
alpha (α)	0.05	Enter values in red
Population Proportion	0.6729	*This is often assumed to be 0.5
Desired Maximum CI Width	0.02	
z	1.96	
Sample Size	2114	<-Rounds up to nearest integer

AQI Bucket	Count	Proportion %	
Good	5,510	6.33%	
Moderate	29,417	33.80%	
Satisfactory	23,636	27.16%	67.29%
Poor	11,493	13.21%	
Very Poor	11,762	13.52%	
Severe	5,207	5.98%	
Grand Total	**87,025**	**100.00%**	

Alpha	0.05
Stdev	131.3243
Sample Size	25000
Sample Average	178.7774
Confidence Interval	95%
Error	1.627883
Lower	177.1496
Upper	180.4053

As the AQI histogram tail again shows a greater potential for extreme outcomes as seen in the city of Ahmedabad. There was also a relatively high kurtosis score and positive skewness number when performing sample data analysis within excel. These ratings backup the notion of a positively skewed AQI distribution in which high outliers need to be taken into consideration.

AQI	
Standard Deviation	131.3243
Sample Variance	17246.08
Kurtosis	8.532544
Skewness	1.930088

Using additional descriptive statistics, a summary chart was created with all decibel day and night data. Sample variance, Kurtosis and skewness were all positive but also all higher when it came to night time decibel levels. This shows slightly more variation at night and a slightly higher risk for any outlier loud noises. With relatively low sample Kurtosis, variance and skewness however, decibel levels throughout these India cities were normally distributed and unimodal. There was also a correlation between day and night noise in the cities,

Day		*Night*	
Standard Deviation	7.407644	Standard Deviation	8.17273
Sample Variance	54.87319	Sample Variance	66.79352
Kurtosis	0.795961	Kurtosis	1.396504
Skewness	0.279136	Skewness	0.723751

as well as a positively high sample covariance. After performing regression statistics on the decibel dataset, there is a strong r value of .88 and a low p-value less than .05. With the positive r value (correlation) it is shown that as a city's decibel level increases during the day, it also increases and stays consistent at night. BigAlData determined a sample covariance between the two variables of 53.47. This all proves a strong relationship for linear association between day and night decibel variables and again backs up the consistency and correlation found without many outliers. There is now no need to segregate these as separate variables when finding a quiet location for India's new Children's Hospital.

Covariance.S	53.47266
Correlation	0.883252

Location	Average of Day	Average of Night
Silence	62.19	56.88
Residential	64.07	58.00
Industrial	66.32	61.48
Commercial	69.04	63.68
Grand Total	**65.83**	**60.43**

City	Average of Day	Average of Night
Delhi	61.79	57.18
Bengaluru	62.88	57.40
Lucknow	65.93	58.97
Mumbai	66.43	62.16
Kolkata	66.47	62.37
Hyderabad	66.77	61.05
Navi Mumbai	69.20	62.53
Chennai	70.17	63.84
Grand Total	**65.83**	**60.43**

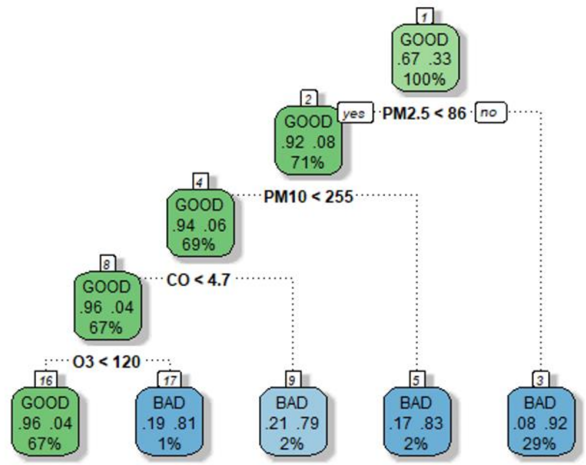
With all of this data gathered, BigAlData believes that AQI and decibel levels may have some mutual exclusivity between each other. This again emphasizes that the most important dataset with the children's health in mind is that of India's AQI. Based on data analyzed sampling and validation findings, BigAlData can now be confident in performing data splits and k-fold cross-validation predictive model building. This will help judge model performance and prevent overfitting prior to predicting what AQI ratings will look like in the future when India's Children's Hospital will be filled with patients.

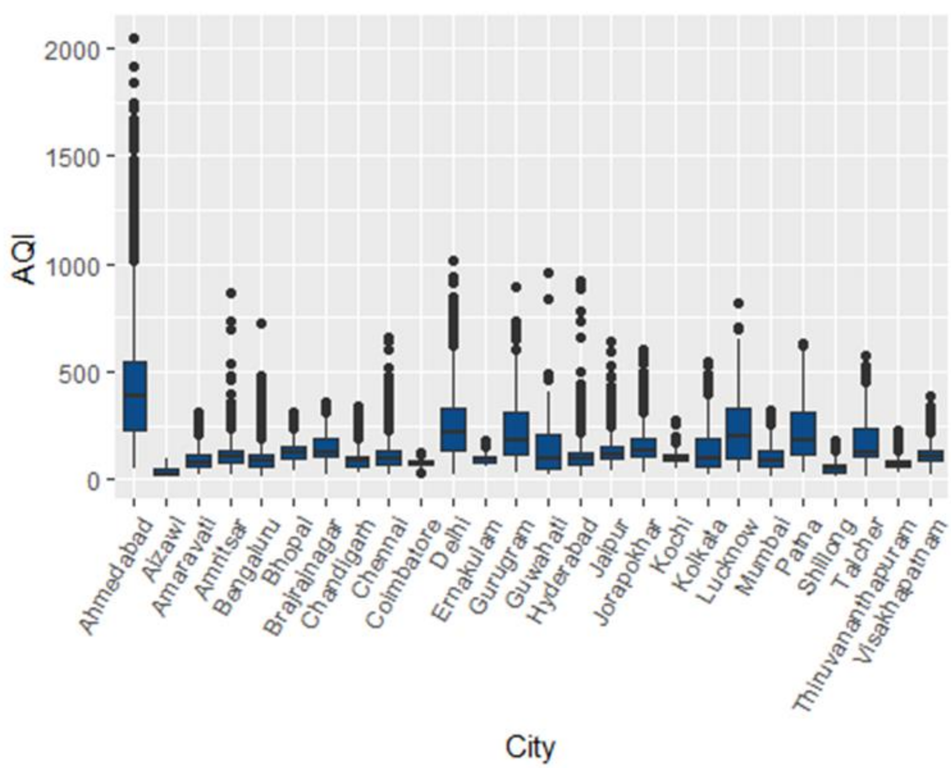
Predictive Modeling and Splits

BigAlData then performed predictive model building on the data sets for India's Air Quality Index and decibel levels to help prove how effective the predictions will be. For realistic results; training and testing splits have been performed, as well as k-fold cross-validation to help simulate the existence of unseen data and prevent overfitting. Due to the large dataset, data is split into a training set consisting of 70% and a testing set with 30% of historical readings. This helps act as a predictive way to show model performance on a portion of data previously collected that the created models have never seen before. K-fold cross-validation is an additional split method used which builds one model for each partition that is held out of the training runs. The more variations a model build can have, the better consistency and prevention of overfitting. This helps come up with feasible and accurate future predictions.

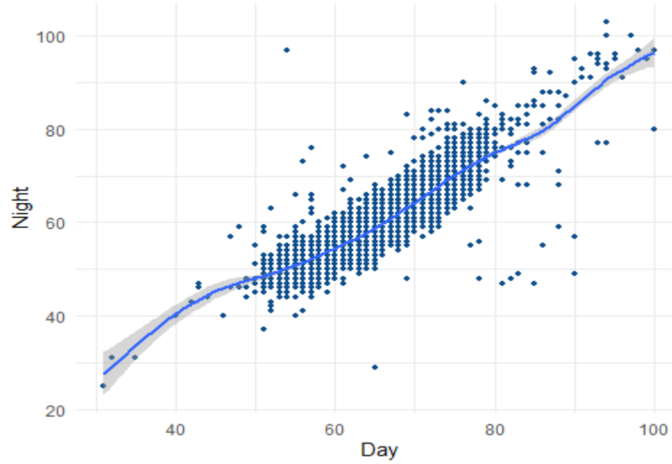
```
Confusion Matrix and Statistics

          Reference
Prediction  GOOD   BAD
      GOOD 16710   638
      BAD    858  7900

               Accuracy : 0.9427
                 95% CI : (0.9398, 0.9455)
    No Information Rate : 0.6729
    P-Value [Acc > NIR] : < 2.2e-16
```

Using a model validation technique within R programming, BigAlData was able to validate testing performance of up to 86.3% accuracy. This was using station locations, dates and readings as variables to help predict AQI. Along with this model validation, a stronger predictive test statistic was produced with 94.27% accuracy using a classification tree technique. Both model methods also incorporated previously explained splitting techniques so that BigAlData could accurately predict future AQI readings without the risk of overfitting to historical data gathered. With the strongest model conducted being the classification tree model, BigAlData was able to accurately predict good (good/moderate/satisfactory) levels 96% of the time and bad (poor/very poor/severe) 90% of the time. The classification tree as shown, split primarily with PM2.5 readings for 71% of those readings leading have good, moderate or satisfactory AQI ratings. A boxplot distribution was also produced within R to highlight potential outlier cities as previously discussed.





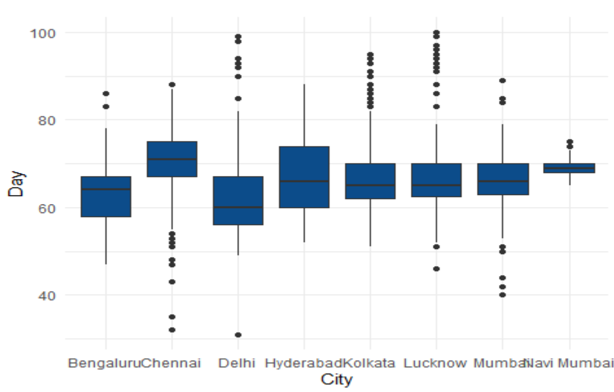
Using predictive statistics of model validation on decibel levels, BigAlData was able to confidently test data gathered from India stations regarding decibel level ratings. As previously confirmed, daytime and nighttime decibel levels had a strong correlation with one another and this correlation graph produced within R again backs up that notion. Although this won't be the main dataset when determining where the Children's Hospital should be built, it still needs to be proven reliable for potential backup predictions. Dataset splits, linear regression and lasso regression were performed within R programming to prevent overfitting of gathered data. It is shown again below that daytime and nighttime accuracy are extremely similar when it came to trusting the model and testing held data. Performance was 85% accurate when predicting daytime decibel levels and 86% accurate when predicting nighttime decibel levels given the historical information provided. This gives a strong sense that decibel levels can be accurately predicted with only a slight amount of error for future levels in specified cities.

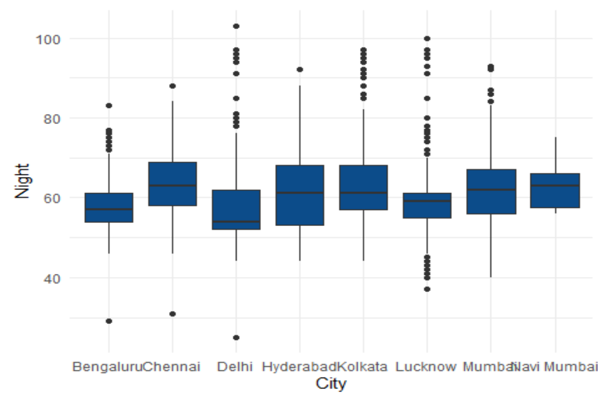



```
last_fit(
  final_lasso,
  Decibel_split) %>%
  collect_metrics()

## # A tibble: 2 x 4
##   .metric .estimator .estimate .config
##   <chr>   <chr>          <dbl> <chr>
## 1 rmse    standard       2.81  Preprocessor1_Model1
## 2 rsq     standard       0.850 Preprocessor1_Model1
```

```
last_fit(
  final_lasso,
  Decibel_split) %>%
  collect_metrics()

## # A tibble: 2 x 4
##   .metric .estimator .estimate .config
##   <chr>   <chr>          <dbl> <chr>
## 1 rmse    standard       3.04  Preprocessor1_Model1
## 2 rsq     standard       0.863 Preprocessor1_Model1
```

Boxplots showing individual city decibel levels were also created within R programming to show variations in decibel levels day and night. This also exemplifies some possible outliers as well as showing Delhi to seemingly have the lowest median of noise during the day as well as the night.

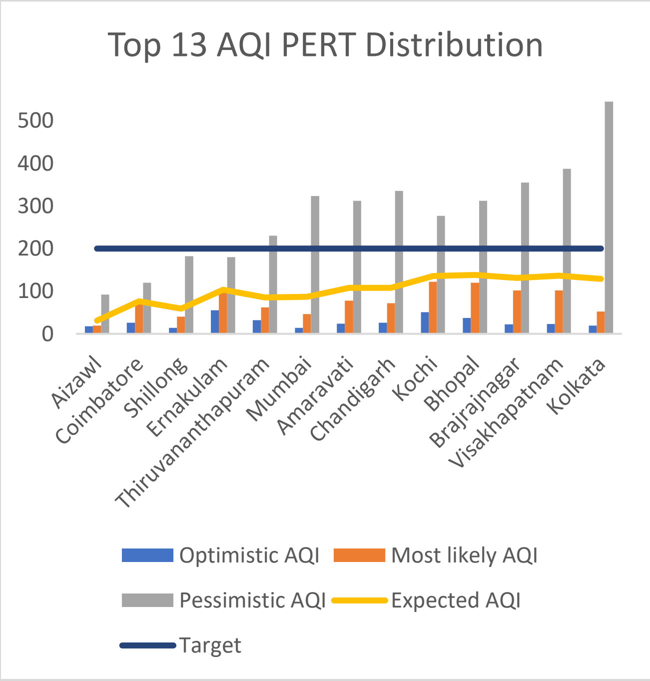




With all of this predictive modeling performed, BigAlData can now confidently believe in accuracy predictions for AQI and decibel levels in the future. This allows BigAlData to correctly predict where India's newest Children's Hospital should be built with children's health in mind when the new facility will be at full capacity. After these model creations in R programing, BigAlData can say that AQI will be accurately predicted with over 94% certainty on air quality and that decibel level ranges will be accurately predicted with over 86% accuracy. Seeing as though the main variable has been highlighted as AQI, BigAlData is extremely pleased with these results. BigAlData will next perform some prescriptive analytical techniques to further support a suggestion for the correct location of the Children's Hospital seeing as though now there is such confidence in predicting environmental aspects that will help children in need.

Prescriptive Model Building

BigAlData lastly focused on prescriptive model building for India's AQI and decibel levels to help optimize which city India's next Children's Hospital should be built. The focus on these prescriptive statistics was geared towards scientifically choosing the best station location that will have the highest future probability of a healthy environment for air quality and noise. To generate these recommendations, a set of constraints first needed to be formed. Once these constraint limits were determined, BigAlData could hone in on leading city candidates that would have the highest probability of low AQI. Once the top half of AQI cities were determined, the secondary dataset of decibel levels could then be viewed and determine if any additional value could be added. Program Evaluation Review Technique (PERT), Triangle Random Variate, and Discrete Random Variables were the main three prescriptive methods used. PERT used optimistic, pessimistic, and most frequent readings to calculate probability of future expected AQI and decibel levels. Triangle Random Variate and Discrete Random Variables use these same factors, as well as sample testing against future randomly generated numbers that provided a view into potential outcomes amongst each city and health rating. Please keep in mind as well that probabilities

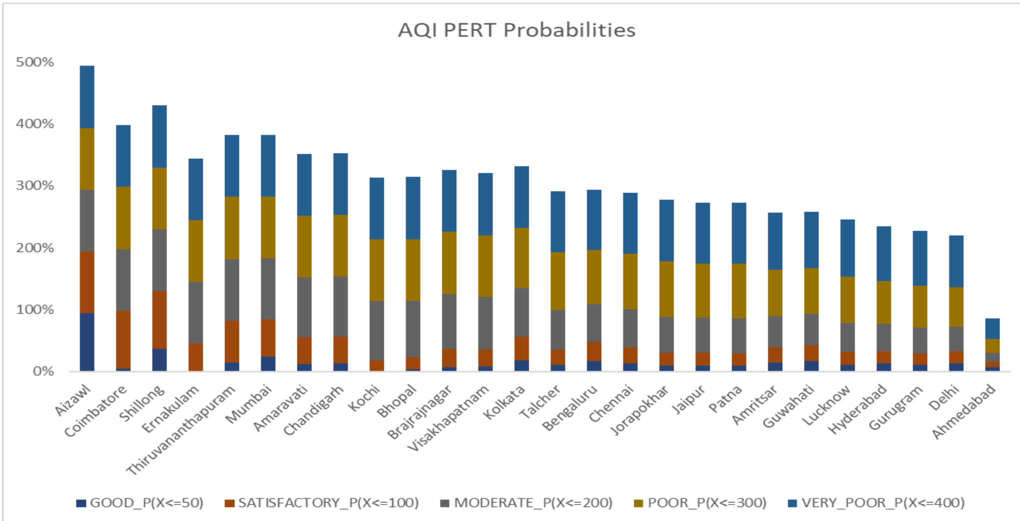


will not perform perfect predictions, as there is no crystal ball to look into the future. However, these scientific findings are as close as one can get to that and will give extremely strong indications on the best fit city to build India's new Children's Hospital.

Using prescriptive analytics techniques of PERT, Triangle Random Variate, and Discrete Random Variables future probabilities of a city having low and healthy AQI can be determined. This was using station locations as variables to help predict future AQI probabilities. AQI constraints of less than 200 were introduced to keep AQI ratings at either good, moderate, or satisfactory levels. The remaining three AQI ratings of poor, very poor, and severe were in turn classified as unhealthy AQI ratings.

PERT distribution allowed BigAlData to cut perspective cities in half by using these calculated expected AQI readings in excel. These expected AQI were derived from using optimistic, pessimistic, and most likely occurring AQI ratings from the dataset. With expected AQI, there was a clear cutoff between top and bottom 13 potential cities where Visakhapatnam had an expected AQI of 136 which was significantly different than the next city of Bengaluru who had an expected AQI of 163. With probabilities added to the PERT grid, there also proved to be a strong probability for future AQI readings on these top 13 cities where they would be either moderate or better (<200 AQI) at least 79% of the time. It is also clear here that the bottom 13 cities have a 1% to 67% potential of recording a day with severe AQI (>400), which we cannot afford to risk with children's health as our primary concern. The top 13 AQI shown left to right in the stacked bar chart will be the best bet at optimal locations for the new hospital.

City	Expected AQI	GOOD_P(X<=50)	SATISFACTORY_P(X<=100)	MODERATE_P(X<=200)	POOR_P(X<=300)	VERY_POOR_P(X<=400)	SEVERE_P(X>400)
Aizawl	**31.00**	**94%**	**100%**	**100%**	**100%**	**100%**	**0%**
Coimbatore	**76.33**	**5%**	**93%**	**100%**	**100%**	**100%**	**0%**
Shillong	**59.33**	**37%**	**93%**	**100%**	**100%**	**100%**	**0%**
Ernakulam	**103.17**	**1%**	**44%**	**100%**	**100%**	**100%**	**0%**
Thiruvananthapuram	**85.00**	**14%**	**68%**	**100%**	**100%**	**100%**	**0%**
Mumbai	**86.83**	**24%**	**60%**	**99%**	**100%**	**100%**	**0%**
Amaravati	**108.00**	**11%**	**43%**	**97%**	**100%**	**100%**	**0%**
Chandigarh	**108.17**	**13%**	**44%**	**96%**	**100%**	**100%**	**0%**
Kochi	**136.00**	**1%**	**17%**	**96%**	**100%**	**100%**	**0%**
Bhopal	**138.17**	**3%**	**20%**	**91%**	**100%**	**100%**	**0%**
Brajrajnagar	**130.83**	**7%**	**29%**	**89%**	**100%**	**100%**	**0%**
Visakhapatnam	**136.33**	**8%**	**27%**	**85%**	**100%**	**100%**	**0%**
Kolkata	**128.67**	**18%**	**37%**	**79%**	**97%**	**100%**	**0%**
Talcher	165.17	11%	24%	65%	93%	99%	1%
Bengaluru	163.83	17%	30%	62%	87%	98%	2%
Chennai	169.00	13%	26%	61%	89%	98%	2%
Jorapokhar	178.50	9%	21%	59%	90%	99%	1%
Jaipur	182.50	9%	21%	57%	88%	98%	2%
Patna	183.33	9%	20%	57%	88%	98%	2%
Amritsar	199.50	15%	24%	50%	76%	92%	8%
Guwahati	199.50	17%	26%	50%	74%	90%	10%
Lucknow	210.83	11%	20%	47%	75%	93%	7%
Hyderabad	223.83	13%	21%	44%	69%	88%	12%
Gurugram	230.33	11%	18%	42%	69%	88%	12%
Delhi	241.33	13%	20%	40%	64%	83%	17%
Ahmedabad	550.83	7%	9%	15%	23%	33%	67%



Random modeling of Triangle Random Variate enabled BigAlData to test sample sizes of potential AQI ratings with the probabilities derived from pessimistic, optimistic, and frequency occurrences. By only viewing Triangle Random Variates of the top 13 cities, focus was given more on the specific desired outcomes of high quality AQI. This random sampling allowed a return of high or low probabilities of healthy AQI and tested those findings. The variation and deviation amongst many cities are still shown here even with probabilities having only small sample testing size of ten per city. This does show how much stronger the top few cities performed against the rest. Aizawl and Coimbatore didn't see any high random sampling larger than 100. This meant that all 20 samples drawn for these two cities consisted of AQI being either good or satisfactory. Although Shillong is expected to draw AQI ratings of moderate or better 100% of the time, there looks to be a wide variance and possibility for an AQI score very close to our 200 max.

Aizawl	***Shillong***	***Coimbatore***	***Thiruvananthapuram***	***Mumbai***
39.61	69.79	64.97	70.94	54.86
18.75279466	164.806522	55.70164276	76.6230872	239.0119648
37.09181187	32.29588603	76.49736483	189.8927043	80.05077905
23.09960729	164.7959596	64.64725656	80.52116854	190.2879511
46.9878925	64.13977216	40.76832242	182.3843403	154.9484966
19.63690172	55.52753327	53.27090565	119.7712943	229.4483431
34.69440442	60.87655203	84.95829494	75.30852421	95.80596229
52.25112491	113.308657	52.84725496	140.2840548	59.36152372
84.51086564	126.6923241	80.23145314	85.91389265	189.4369201
46.56469631	40.86539125	91.47730136	180.3743187	86.34946119
Ernakulam	***Amaraviti***	***Chandigarh***	***Kolkata***	***Brajrajnagar***
69.69	106.92	101.78	85.31	111.63
75.50527546	253.3959959	98.29186331	239.9718584	51.471704
94.38348994	70.19748003	248.9279112	313.2293104	48.24407263
83.80091078	230.9400714	204.6486925	269.3000994	260.8628603
95.44279979	165.1633562	91.26455979	237.5783577	178.5150499
107.249145	159.4616182	91.47868266	310.5924275	241.6489075
102.7484129	59.08373911	152.4204583	335.9666869	201.9404779
157.7876024	84.49048453	310.5963365	324.260641	285.754792
153.3164262	85.49132089	147.9562259	151.0016711	95.01858909
131.7828216	174.9704298	78.17299771	62.03782986	77.45702185
	Kochi	***Visakhapatnam***	***Bhopal***	
	112.71	167.19	116.80	
	139.431056	58.47327751	235.9865142	
	150.1129731	342.8167929	109.9253076	
	241.6375773	299.5653181	125.503938	
	90.28752714	214.7295215	111.8350567	
	268.5107942	106.4387337	217.7057638	
	105.8628045	39.08775307	137.9457735	
	240.5684038	160.5786385	265.320342	
	163.8178423	60.51243237	140.491724	
	125.032355	245.0949693	184.8940395	

Lastly with the AQI data, BigAlData performed an additional random sampling with all data and two subsets (top/bottom 13) showing AQI rating probabilities. This random sampling proves AQI readings with top 13 cities will have less than a 10% chance of obtaining a poor, very poor or severe AQI rating and the riskiness of our bottom 13 cities have a combined 36% chance of returning an AQI of poor, very poor or severe. Although only using an extremely small sample size of ten random numbers, the differentiation of top 13 cities is shown from the rest of the field. This gives extreme confidence in BigAlData's findings and ability to predict a healthy future AQI amongst the selected top half of cities being researched.

Top 13

Row Labels	Count of AQI_Bucket	Lower	Upper	Demand	Random Number	
Good	16.77%	0	0.167736	Good	0.285761621	Satisfactory
Satisfactory	40.52%	0.167736	0.572984	Satisfactory	0.74659698	Moderate
Moderate	31.44%	0.572984	0.887337	Moderate	0.67376822	Moderate
Poor	8.19%	0.887337	0.969215	Poor	0.973727427	Very Poor
Very Poor	2.84%	0.969215	0.997644	Very Poor	0.956773603	Poor
Severe	0.24%	0.997644	1	Severe	0.603398072	Moderate
Grand Total	**100.00%**				0.629032208	Moderate
					0.956179566	Poor
					0.294160279	Satisfactory
					0.184128464	Satisfactory

All

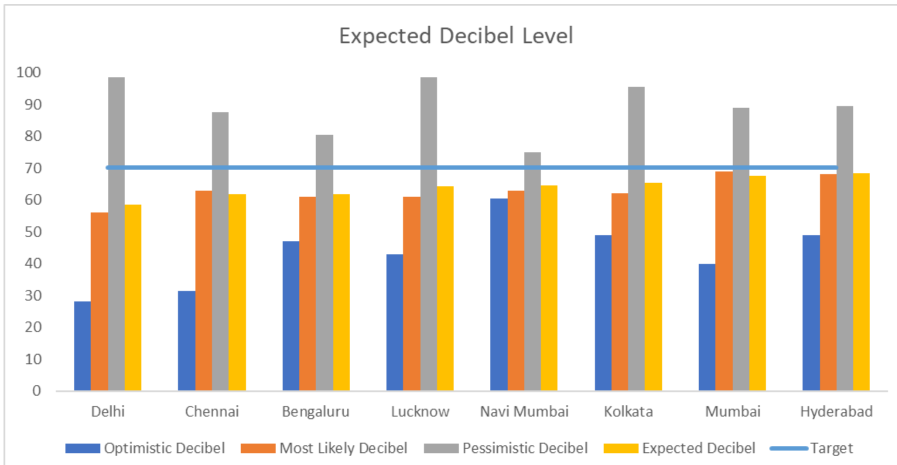
Row Labels	Count of AQI_Bucket	Lower	Upper	Demand	Random Number	
Good	6.33%	0	0.063315	Good	0.285761621	Satisfactory
Satisfactory	27.16%	0.063315	0.334915	Satisfactory	0.74659698	Poor
Moderate	33.80%	0.334915	0.672945	Moderate	0.67376822	Poor
Poor	13.21%	0.672945	0.80501	Poor	0.973727427	Severe
Very Poor	13.52%	0.80501	0.940167	Very Poor	0.956773603	Severe
Severe	5.98%	0.940167	1	Severe	0.603398072	Moderate
Grand Total	**100.00%**				0.629032208	Moderate
					0.956179566	Severe
					0.294160279	Satisfactory
					0.184128464	Satisfactory

Bottom 13

Row Labels	Count of AQI_Bucket	Lower	Upper	Demand	Random Number	
Good	4.61%	0	0.046109	Good	0.285761621	Satisfactory
Satisfactory	24.96%	0.046109	0.295688	Satisfactory	0.74659698	Poor
Moderate	34.19%	0.295688	0.637618	Moderate	0.67376822	Poor
Poor	14.03%	0.637618	0.777953	Poor	0.973727427	Severe
Very Poor	15.27%	0.777953	0.930696	Very Poor	0.956773603	Severe
Severe	6.93%	0.930696	1	Severe	0.603398072	Moderate
Grand Total	**100.00%**				0.629032208	Moderate
					0.956179566	Severe
					0.294160279	Satisfactory
					0.184128464	Satisfactory

City	Prob < 70 Decibel
Navi Mumbai	99%
Bengaluru	93%
Delhi	84%
Chennai	81%
Lucknow	73%
Kolkata	72%
Mumbai	62%
Hyderabad	59%

It was previously confirmed that daytime and nighttime decibel levels had a strong correlation with one another so a simple average was taken based on those statistics. Based on the Hearing Health Foundation's research findings, BigAlData added a constraint that decibels needed to be 70 or under to remain safe for the children's ears. https://hearinghealthfoundation.org/decibel-levels With that in mind, probabilities were determined using PERT analysis to show the future odds of a city having at most 70 decibels during both day and night. Per the chart, there is no guarantee any city would score 100%, however Navi Mumbai, Bengaluru, Delhi, and Chennai all had a probability of over 80% that they would have safe decibel levels at 70 or less. The problem here however is that none of these four cities were previously shown as being top 13 from an AQI perspective. The only two cities with decibel level data that were in the top 13 AQI are Kolkata and Mumbai. Each of those cities did have expected decibel levels of 65 and 68 respectively which are safe, but they have probabilities of 28% and 38% respectively with the potential for producing unsafe hearing decibel levels at one point or another. This data will again be used as supporting statistics with expected decibel levels all remaining below 70, but are using this as an inferior dataset alongside AQI.



Using random testing with Triangle Random Variate, BigAlData can exemplify the possible decibel variability among each of the eight cities with decibel level data. Although this is only a small sample size of 10 random checks based on probabilities, each city shows at one point or another a time in which decibel levels reached over 70. Kolkata and Mumbai had several occurrences where decibels reached close to 90 or higher which was the highest among any city shown in our sample. With this in mind, BigAlData believes it is safe to eliminate these two cities from the top 13 based on relatively high probability that each location will experience times during the day with higher-than-normal decibel level.

Navi Mumbai		Bengaluru		Delhi		Chennai
62		70		69		36
71		73		54		65
67		55		58		36
64		60		36		68
67		71		67		52
66		72		81		73
67		61		77		66
73		54		39		55
64		58		62		53
68		67		73		73
Lucknow		**Kolkata**		**Mumbai**		**Hyderabad**
56		76		73		57
53		63		88		72
79		61		65		60
71		91		62		63
71		72		67		83
55		64		84		67
64		59		69		84
63		66		65		63
59		63		70		56
72		65		74		72

Using random testing with Discrete Random Variables, BigAlData can show decibel level spikes using charts with all cities included or just our predetermined top 13 based on AQI. During the day, 73% of all cities with data gathered had safe decibel readings while at night, nearly 89% of decibel levels stayed safe. However, even with a small sample size, there are multiple occurrences day and night in which unsafe decibel levels are reached. With this knowledge of inconsistent decibel level variance, our idea of AQI ratings being more important and carrying the majority of influence on hospital location decision remains intact.

City	(All)							
Day	**Count of Day2**	Lower	Upper	Demand	Random Number Day		Random Number Night	
Not Safe	27.09%	0.00	0.27	Not Safe	0.48	Safe	0.97	Safe
Safe	72.91%	0.27	1.00	Safe	0.55	Safe	0.60	Safe
Grand Total	**100.00%**				0.17	Not Safe	0.64	Safe
					0.82	Safe	0.07	Not Safe
City	(All)				0.47	Safe	0.45	Safe
					0.18	Not Safe	0.40	Safe
Night	**Count of Night2**	Lower	Upper	Demand	0.37	Safe	0.77	Safe
Not Safe	11.33%	0.00	0.11	Not Safe	0.67	Safe	0.06	Not Safe
Safe	88.67%	0.11	1.00	Safe	0.16	Not Safe	0.50	Safe
Grand Total	**100.00%**				0.89	Safe	0.03	Not Safe

City	(Multiple Items)							
Day	**Count of Day2**	***Lower***	***Upper***	***Demand***	Random Number Day		Random Number Night	
Not Safe	23.19%	0.00	0.23	Not Safe	0.82	Safe	0.64	Safe
Safe	76.81%	0.23	1.00	Safe	0.02	Not Safe	0.78	Safe
Grand Total	**100.00%**				0.68	Safe	0.42	Safe
					0.76	Safe	0.85	Safe
City	(Multiple Items)				0.76	Safe	0.81	Safe
					0.57	Safe	0.19	Safe
Night	**Count of Night2**	***Lower***	***Upper***	***Demand***	0.89	Safe	0.95	Safe
Not Safe	12.12%	0.00	0.12	Not Safe	0.18	Not Safe	0.00	Not Safe
Safe	87.88%	0.12	1.00	Safe	0.53	Safe	0.04	Not Safe
Grand Total	**100.00%**				0.30	Safe	0.78	Safe

Recommendation

All descriptive, predictive and prescriptive modeling has now been performed and BigAlData can confidently predict probabilities of consistent AQI and decibel levels for India's cities in the future. With the support of statistical probabilities and randomly testing sample techniques, we can predict with a probability of at least 99.9% and an accuracy confidence level of 94.3% where India's newest Children's Hospital should be built. This is prediction supports our goal of keeping children's health in mind and an AQI that will always be either moderate, satisfactory, or good. AQI is by far the most important variable here and will be taken as our sole consideration when choosing a destination city.

Seeing as though a pattern was previously notated with AQI spikes during the months of October through March and we don't have more than five months' worth of AQI readings in Aizawl,

City	Expected AQI	GOOD_P(X<=50)	SATISFACTORY_P(X<=100)	MODERATE_P(X<=200)
Aizawl	**31.00**	**94%**	**100%**	**100.0000%**
Coimbatore	**76.33**	**5%**	**93%**	**100.0000%**
Shillong	**59.33**	**37%**	**93%**	**100.0000%**
Ernakulam	**103.17**	**1%**	**44%**	**99.9998%**
Thiruvananthapuram	**85.00**	**14%**	**68%**	**99.9754%**

BigAlData has chosen to eliminate Aizawl as a potential city. Although Aizawl has the lowest numbers, the fact that not enough historical data may have been gathered to account for potential outliers or swings add risk that BigAlData isn't comfortable passing along to the children of India. Coimbatore and Shillong do produce data back to 2019 so we will use those remaining top two cities as our main focus.

With a much lower expected future AQI rating of 59.33 and a probability of 37% good AQI as well as 93% satisfactory and 100% moderate, BigAlData is recommending to India's healthcare system that the new Children's Hospital be built in the city of Shillong. Shillong with all analytical findings referenced, is the location in India that will provide the most environmentally safe place for the new Children's Hospital to be built. With this path being chosen, there is supreme confidence from BigAlData that these children in India will have a much higher quality of life and their young breaths on earth will remain pure while visiting or staying in their beautiful and clean new hospital they can now call home.