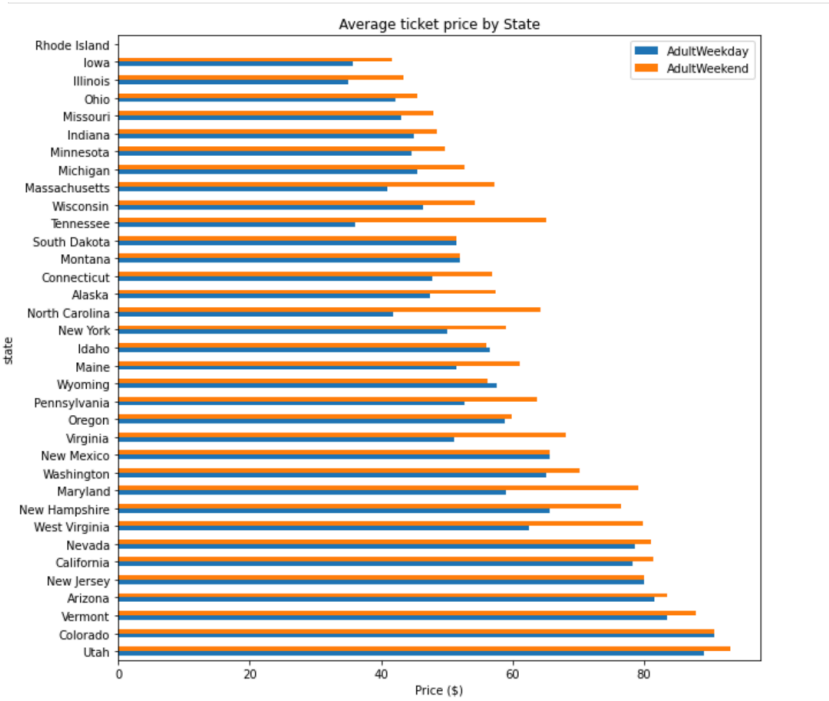
Guided Capstone Project Report

Big Mountain resort in Montana is dealing with a capitalization issue where they cannot determine whether to increase the ticket prices based on the facilities they have by comparing with other resorts or cut the budgets. After identifying the problem, I started analyzing the data to prepare it and model it for presentation. I followed several steps to complete the data analysis including data wrangling, exploratory data analysis, pre-processing and training data development and modeling.

At first there were 330 Entries and 27 columns. Some of the columns that had missing values; fastEight was the column with most missing values with 50.30%. Crystal Mountain resort appeared twice while checking for duplicate values using value_counts() method. However, later it was found to have two different regions and states with the same resort name . From the distributions of resorts graph New York has the most resorts. The targeted resort is in Montana which was the 13th. The resort with an incredibly large skiable terrain area was the Silverton Mountain in Colorado. Weekend tickets had more demand than weekday tickets. Some of the rows did not have values for AdultWeekday and AdultWeekend so they were dropped.



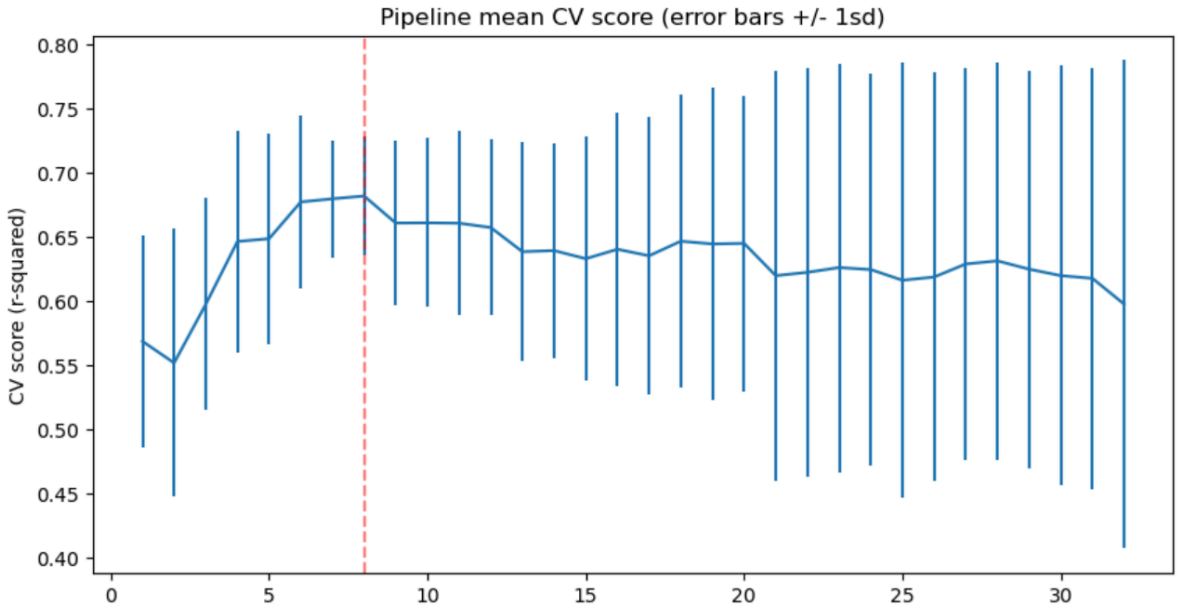
Colorado has the most resorts; 22 with highest total skiable area ac of 43862.0

When it comes to the total state area, Alaska is the highest one with the value of 665384. Our data home state Montana comes in third place. California dominates the state population figures despite coming in second behind Alaska in size (by a long way). The resort's state of Montana was in the top five for size, but doesn't figure in the most populous states. New York has the highest number of resorts with 33 resorts. Montana is one of the top 5 states that has the greatest skiable area. New York dominates the area of skiing available at night. Looking at the top five in general, they are all the more northerly states.

	state	resorts_per_state	state_total_skiable_area_ac	state_total_days_open	state_total_terrain_parks	state_total_nightskiing_ac
0	Alaska	3	2280.0	345.0	4.0	580.0
1	Arizona	2	1577.0	237.0	6.0	80.0
2	California	21	25948.0	2738.0	81.0	587.0
3	Colorado	22	43682.0	3258.0	74.0	428.0
4	Connecticut	5	358.0	353.0	10.0	256.0

Using the 70/30 train/test split we got the size (193.2, 82.8) Used sklearn's DummyRegressor to figure out the mean, coefficient of determination, mean absolute error and mean squared error of the train and test. Imputed missing values with mean and median. Created pipes for linear regression and random regression; random regression had better performance. Saved the best model object from the pipeline.

```
#Code task 20#
#Assign the value of k from the above dict of `best_params_` and assign it to `best_k`
best_k = lr_grid_cv.best_params_['selectkbest__k']
plt.subplots(figsize=(10, 5))
plt.errorbar(cv_k, score_mean, yerr=score_std)
plt.axvline(x=best_k, c='r', ls='--', alpha=.5)
plt.xlabel('k')
plt.ylabel('CV score (r-squared)')
plt.title('Pipeline mean CV score (error bars +/- 1sd)');
```



The goal is to adjust the price of the Big Mountain resort's tickets. So the Big Mountain data was excluded from the dataset to avoid biased results. After calculating the Big Mountain resort's expected ticket price, it turns out there is room to increase the price since the modeled price is $95.87 and the actual price is $81. The business has shortlisted some options like permanently closing down up to 10 of the least used runs. This doesn't impact any other resort statistics, increasing the vertical drop by adding a run to a point 150 feet lower down but requiring the installation of an additional chair lift to bring skiers back up, without additional snow making coverage , increasing the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres. The expected number of visitors over the season is 350,000 and, on average, visitors ski for five days.

```
plot_compare('Snow Making_ac', 'Area covered by snow makers (acres)')
```