The Big Mountain Pricing Strategy

LINA GAO

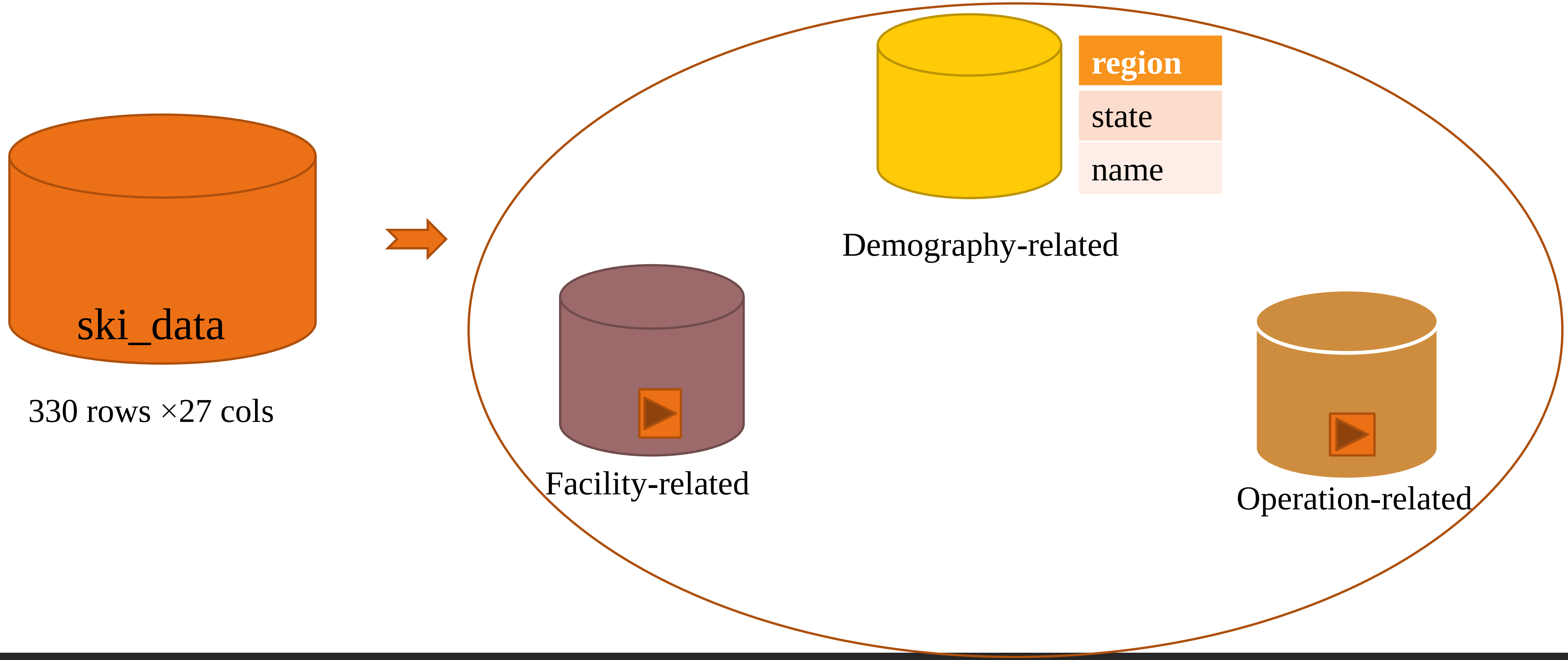
Introduction

- The Big Mountain Resort, a ski resort located in Montana, offers spectacular views of Glacier National Park and Flathead National Forest. ~ 350,000 people ski or snowboard at Big Mountain each year. It provides access to 105 trails for skiers and riders of all levels and abilities.
- It has recently installed a chair lift, which increases the operating costs by $1,540,000 this season.
- The resort's executive team needs suggestions to develop a strategy to cover this cost, either by charging a premium fee, cutting underused facilities, or other scenarios.

Problem identification

- Is the current ticket price of Big Mountain Resort reasonable in the ski resort market with the facilities they provided for the customers?
- If it is lower than its peers, how much can we increase the ticket price without a significant effect on the customer numbers?
- Which facility is the most valuable one to attract customers to visit this resort?
- Is any facility of the resort underused and can be eliminated?

Dataset Overview



Dataset Overview

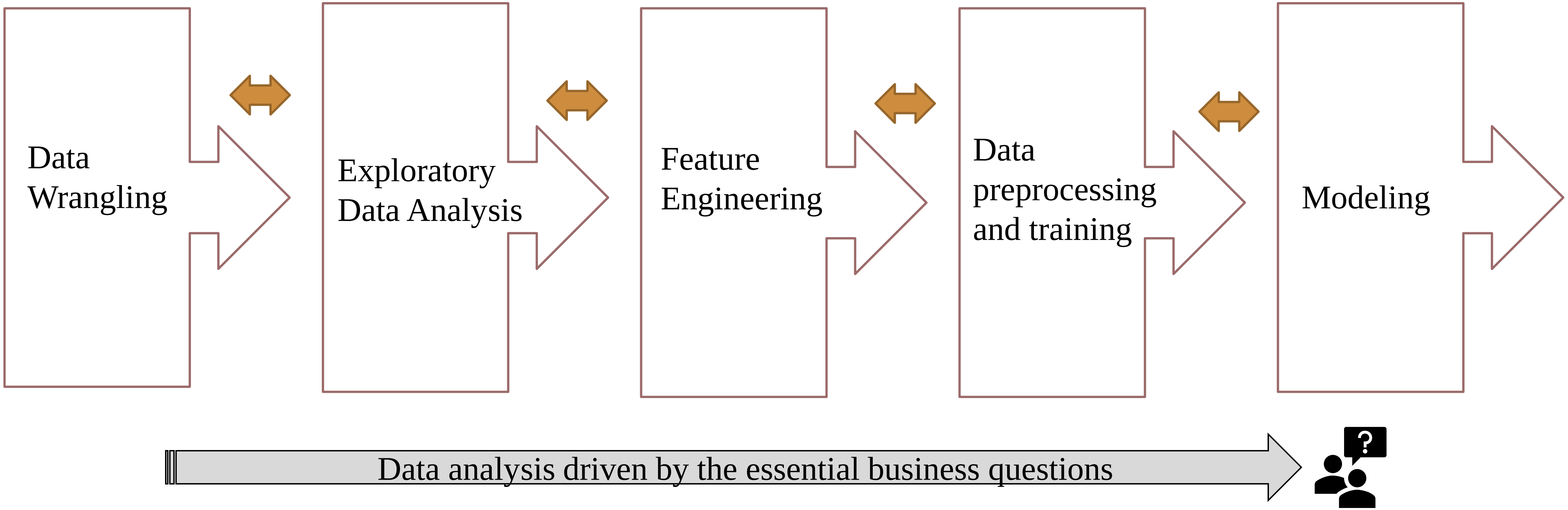
facility-related columns

Column	
Summit_elev	LongestRun_mi
Vertical drop	SkiableTerrain_ac
Base_elev	Snow Making_ac
Surface	averageSnowfall
TerrainParks	NightSkiing_ac

Operation-related columns

Column	
Trams	Runs
fastEight	daysOpenLastYear
fastSixes	yearsOpen
fastQuads	*AdultWeekday*
Triple	AdultWeekend
Double	projectedDaysOpen
Total chairs	

Project workflow



Data wrangling & feature engineering

- ❑

I. Is the dataset clean enough for exploratory data analysis
II. Which feature is out best choice as the target variable?
III. Are there enough proper features for the development of the pricing model ?

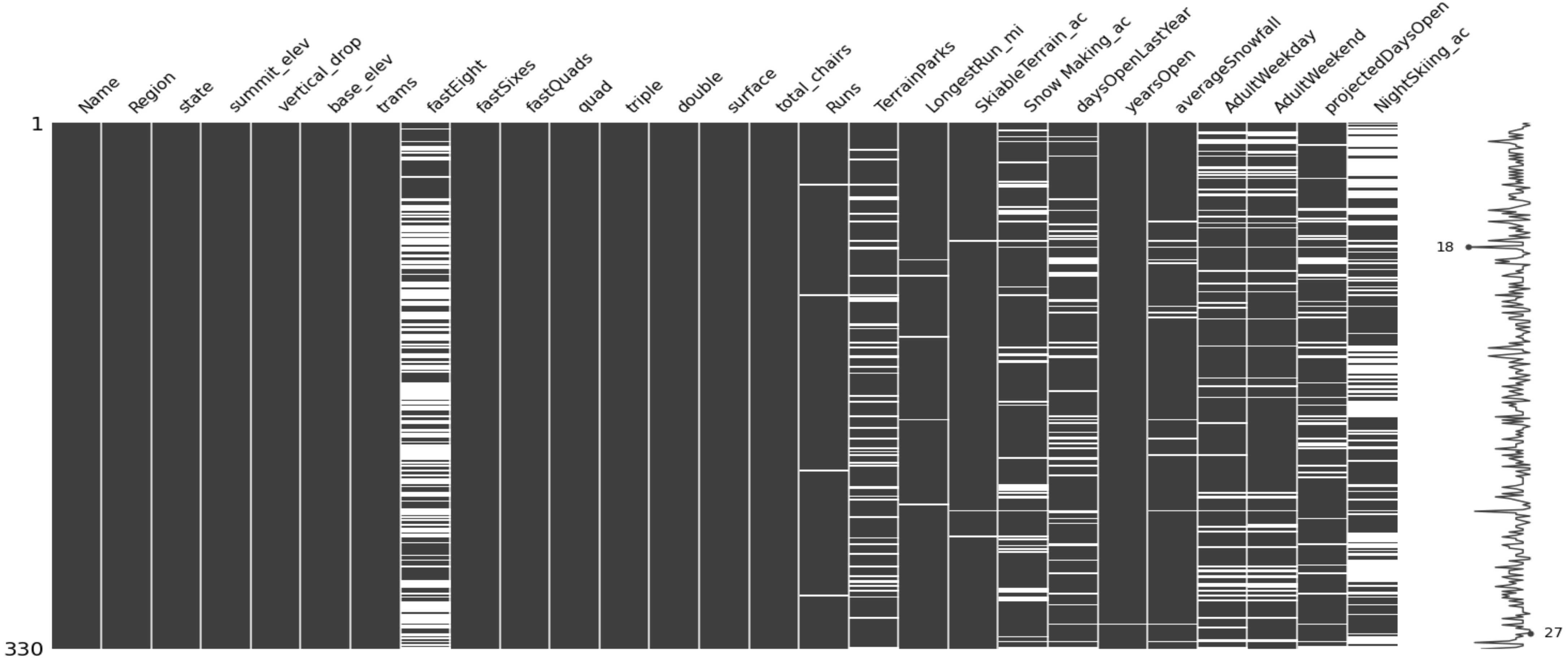
- ❑

Pandas, numpy, matplotlib, seaborn

- ❑

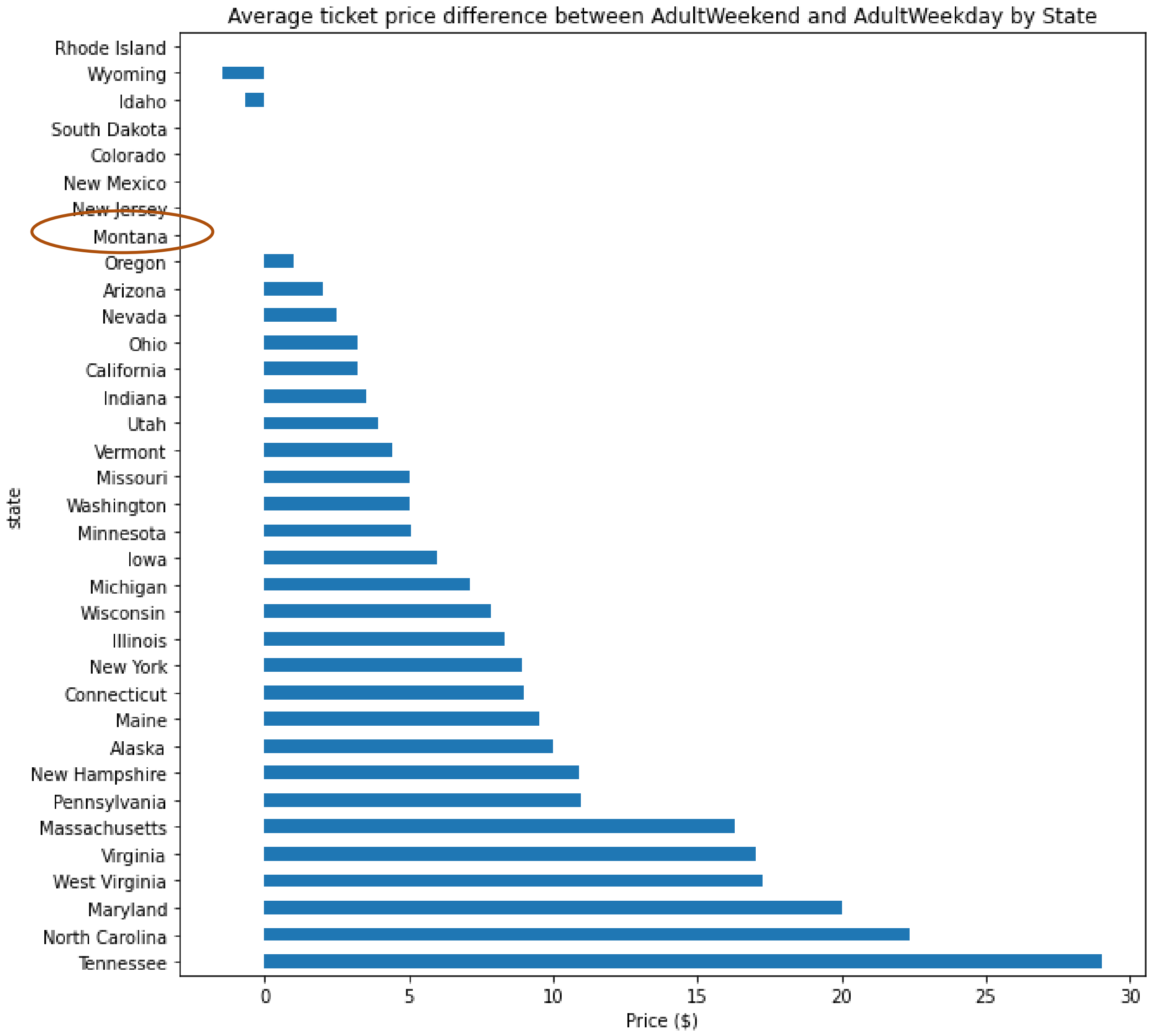
I. Cleaned the missing data, deleted negligible features, corrected suspicious data
II. AdultWeekend is chosen as the proper target variable for the following analysis
III. Added State summary for this project and finalized the data for EDA analysis

Data wrangling & feature engineering
Missing Data Distribution



The Target Variable

- AdultWeekend: 15.45% missing data
- AdultWeekday: 16.36% missing data.



Data wrangling & feature engineering

The novel state-summary data set

- The shape of the data set: (35 rows, 8 cols)
- The variables in this dataset:

The variables in state-summary dataset		
State	State_total_days_open	State_population
Resorts_per_state	State_total_terrain_parks	State_area_sq_miles
State_total_skiable_area_ac	State_total_nightskiing_ac	

Data wrangling & feature engineering
The cleaned ski_data

- The shape of the data set: (277 rows, 25 cols)
- The target variable: AdultWeekend
- The predictors:

  - Geography related variables: no change
  - Facility related variables: delete two variables, *AdultWeekday* and *fastEight*
  - Operation related variables: no change

Exploratory Data Analysis & feature engineering

I. What does the ski-resort market for Montana look like compared with the other states?
II. When we develop the pricing model, should we use all available data or divide the data into Montana and the other states?

* pandas * numpy * matplotlib * seaborn * sklearn

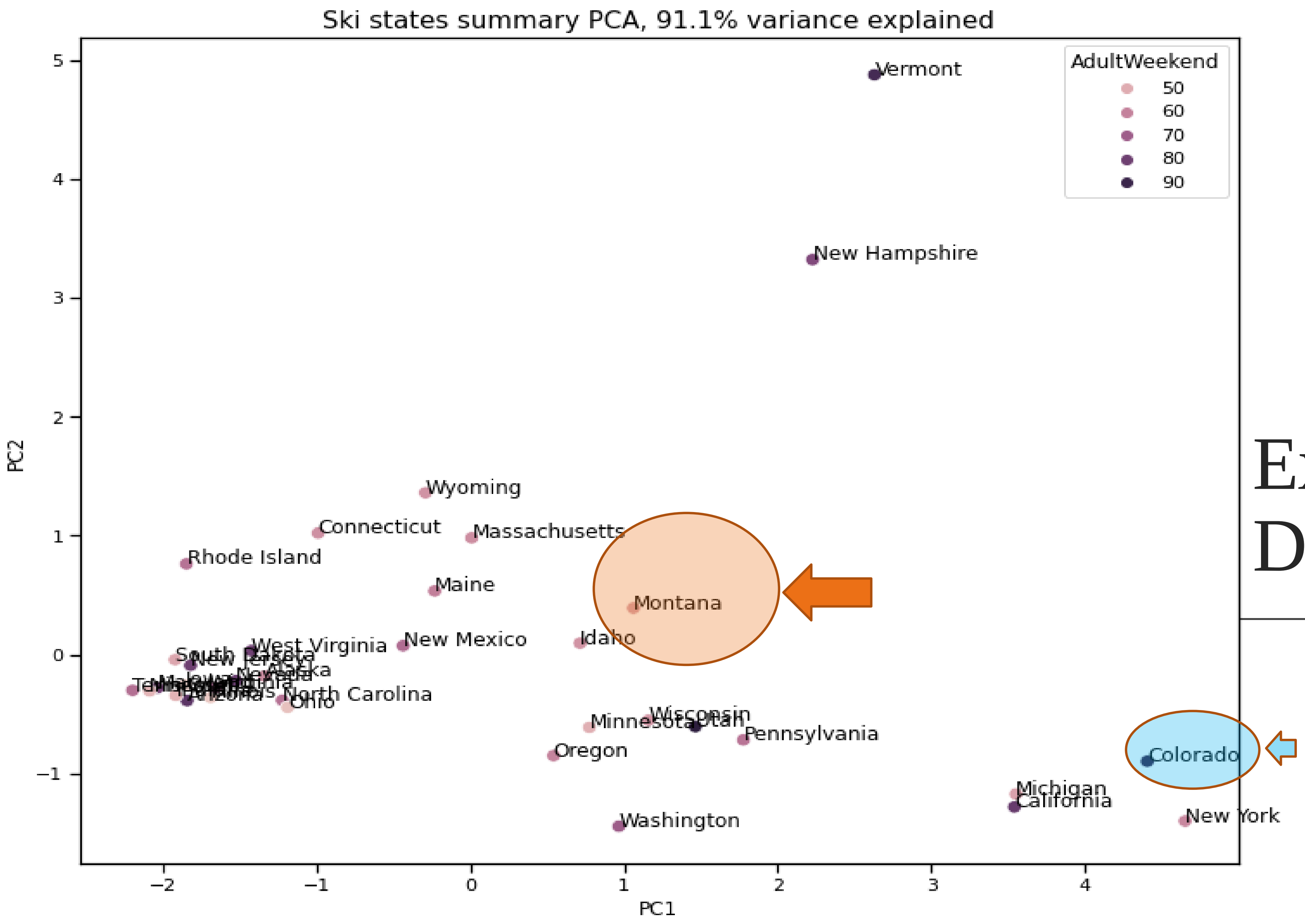
I. Montana is among the top 5 state with the largest state area, skiable area, and the highest number of resorts per 100k capita
II. We can use all available data to develop the pricing model after we add relevant statewide aggregated statistic predictors into ski_data

Exploratory Data Analysis

The top five states in each statewide variables

state	state_area_sq_miles	state	state_population	state	resorts_per_state	state	state_total_skiable_area_ac
Alaska	665384	California	39512223	New York	33	Colorado	43682
California	163695	New York	19453561	Michigan	28	Utah	30508
Montana	147040	Pennsylvania	12801989	Colorado	22	California	25948
New Mexico	121590	Illinois	12671821	California	21	Montana	21410

state	state_total_nightskiing_ac	state	state_total_days_open	state	resorts_per_100kcapita	state	resorts_per_100ksq_mile
New York	2836	Colorado	3258	Vermont	2.40388853	New Hampshire	171.1412985
Washington	1997	California	2738	Wyoming	1.382267922	Vermont	155.9900166
Michigan	1946	Michigan	2389	New Hampshire	1.176720641	Massachusetts	104.2258859
Pennsylvania	1528	New York	2384	Montana	1.122777602	Connecticut	90.20386073

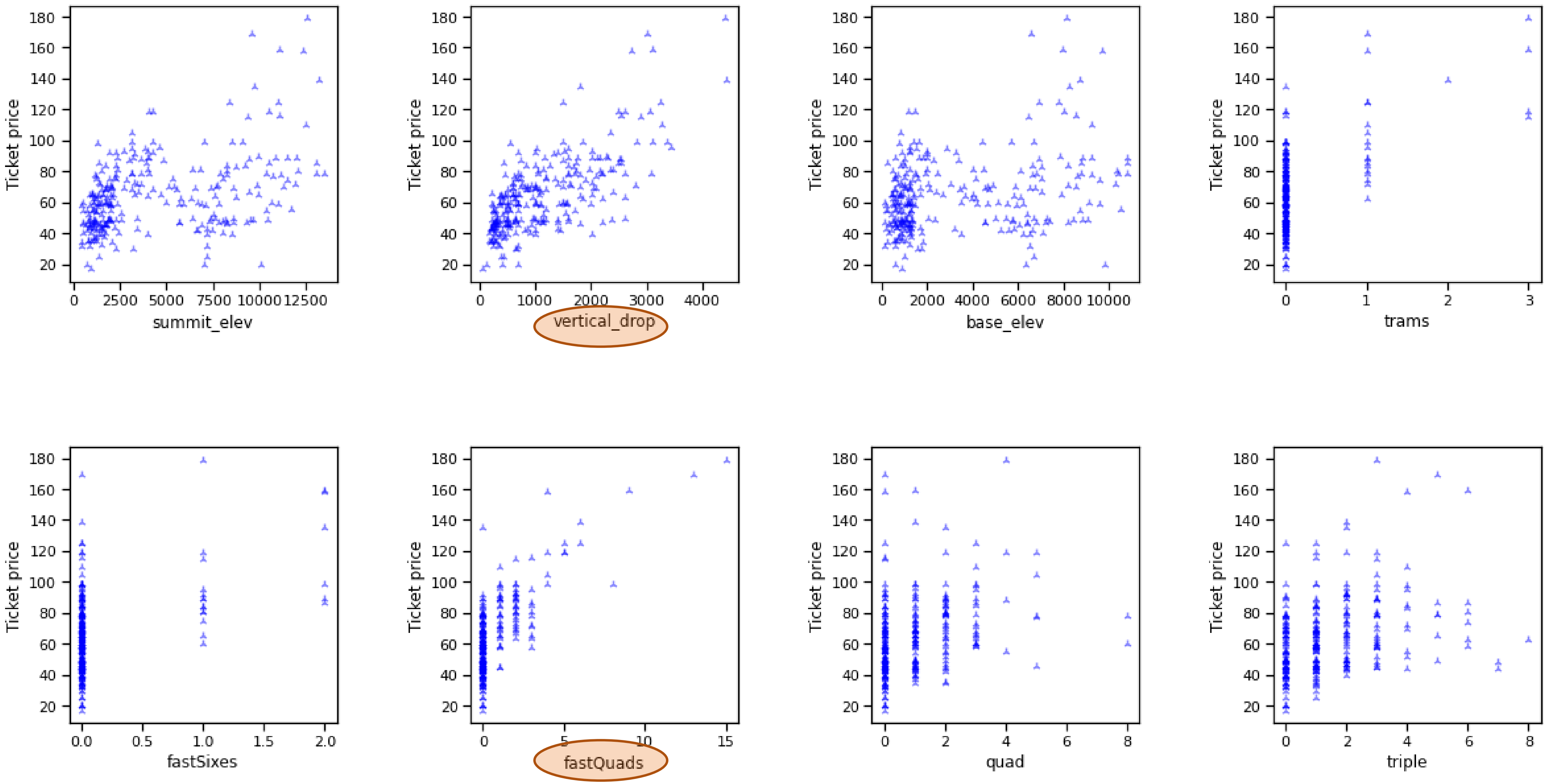


Exploratory Data Analysis

Exploratory Data Analysis & Feature Engineering

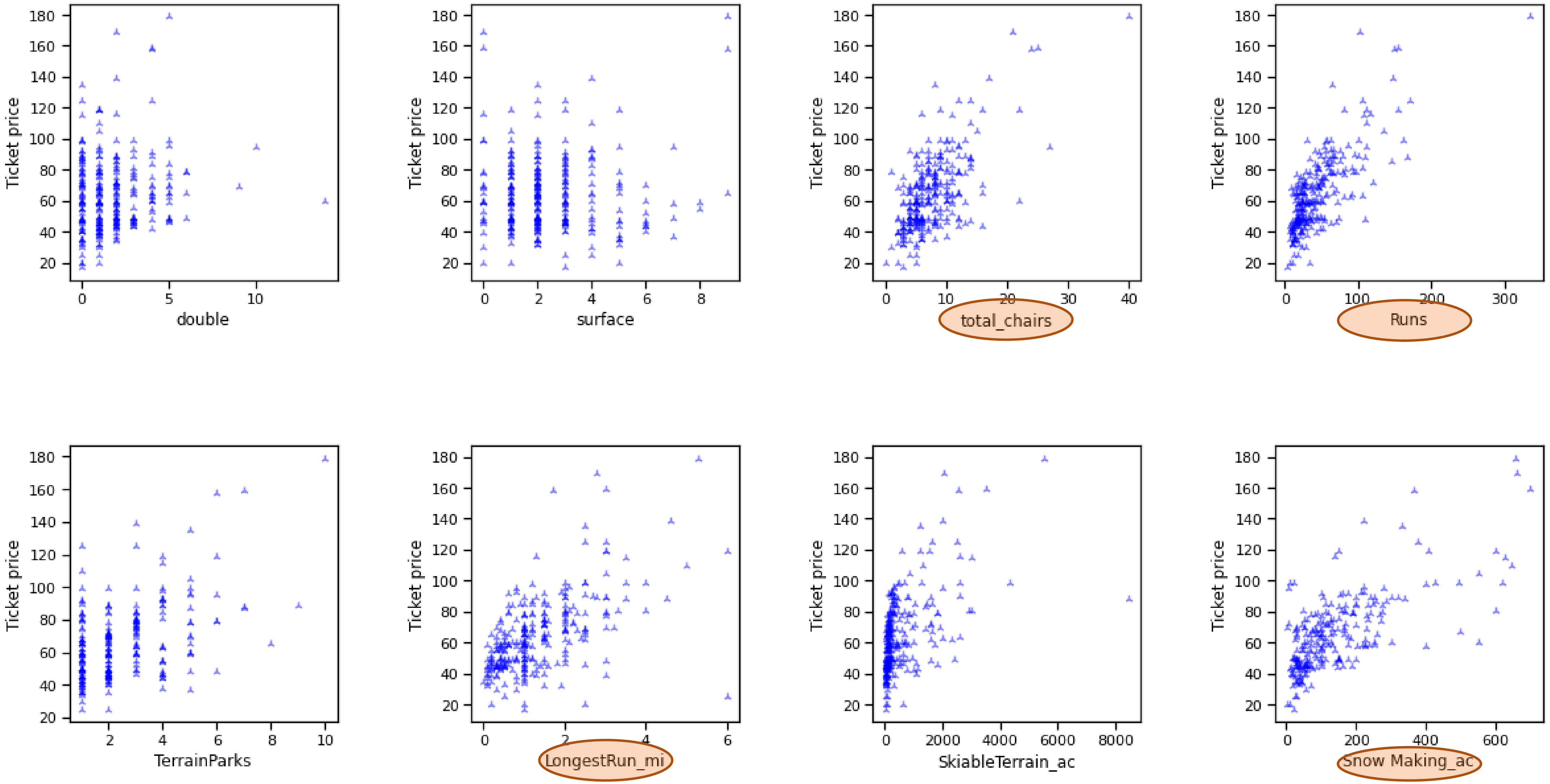
Feature Engineering

Novel features in ski_data	
Resorts_per_state	Resort_night_skiing_state_ratio
Resorts_per_100kcapita	Total_chairs_runs_ratio
Resort_skiable_area_ac_state_ratio	Total_chairs_skiable_ratio
Resort_days_open_state_ratio	FastQuads_runs_ratio
Resort_terrain_park_state_ratio	FastQuads_skiable_ratio



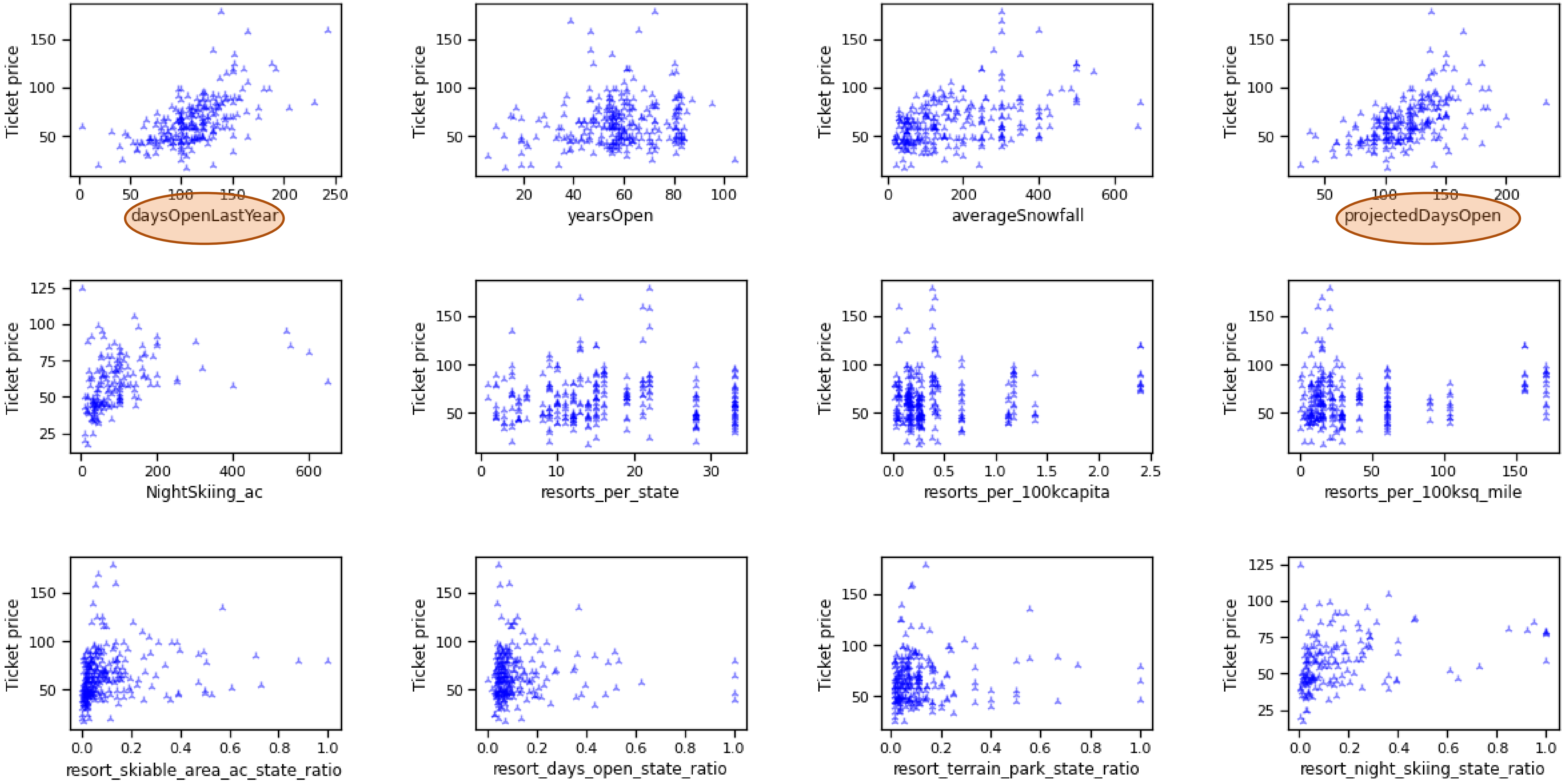
Exploratory Data Analysis & Feature Engineering

Feature Importance



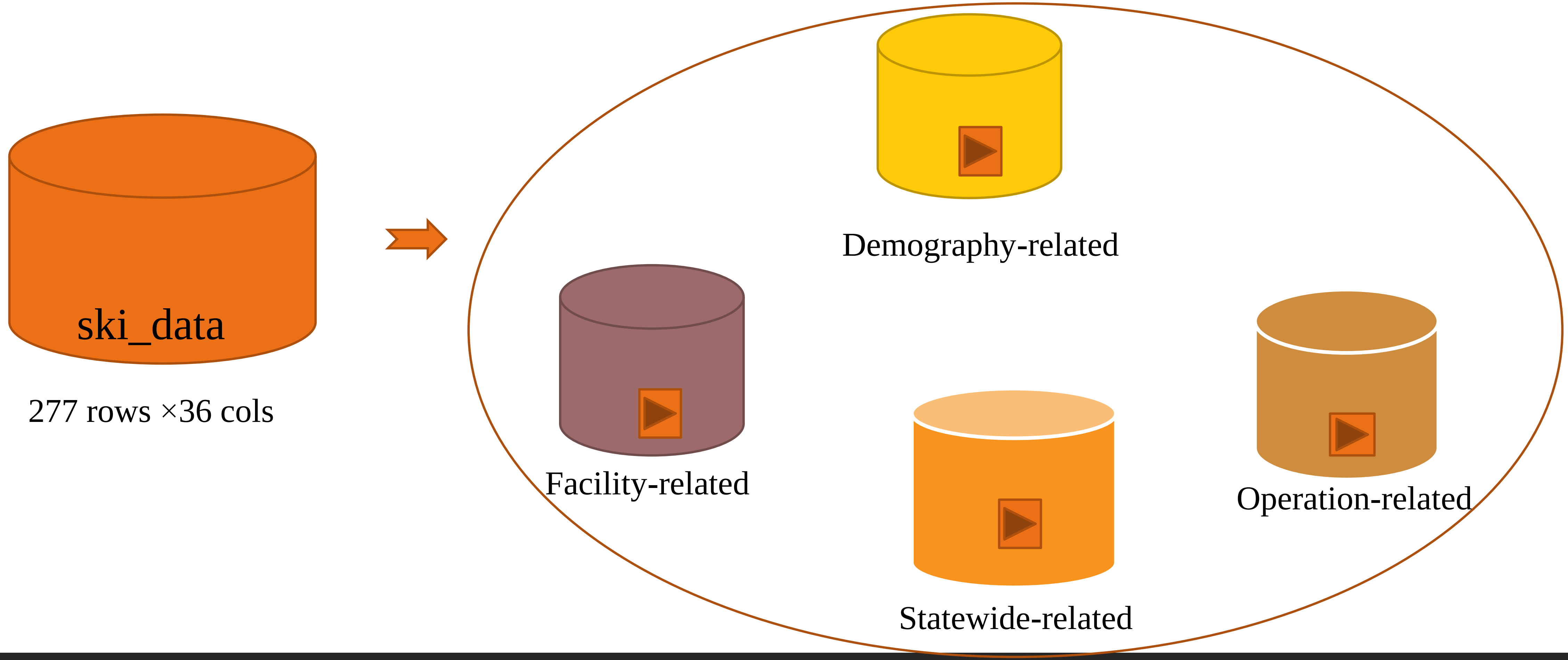
Exploratory Data Analysis & Feature Engineering

Feature Importance



Exploratory Data Analysis & Feature Engineering
Feature Importance

Final Dataset Overview



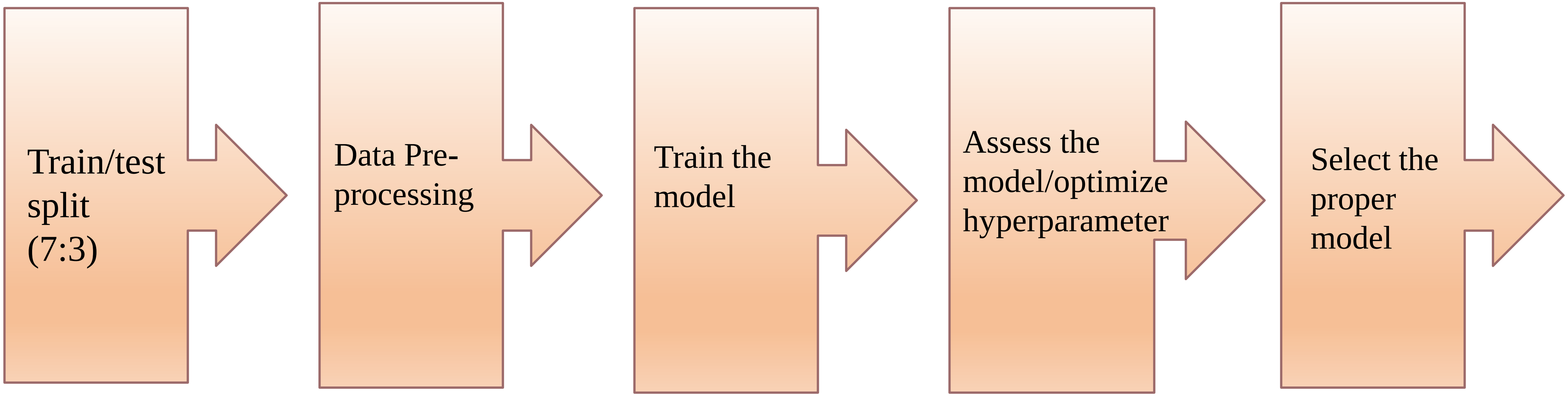
Data Pre-processing and Training

- I. Primary question: How can we develop a good predictive pricing model?
- II. Minor questions:
  1. Which imputation technique should we use for filling in the missing data, mean or median?
  2. How do we train the model?
  3. How can we find the best hyperparameters?
  4. How can we evaluate the model?

  ….

* pandas * numpy * sklearn * matplotlib * seaborn * pickle

Data Pre-processing and Training
sklearn pipeline



X: ski_data-y-demography variables
y: 'AdultWeekend'

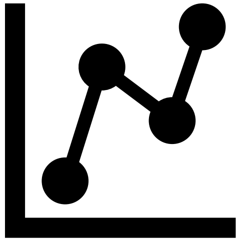
Impute missing values
Scale the data

Optimization technique:
RandomSearchCV
GridSearchCV
Bayesian optimization

Regression model metrics:
R-squared
Mean absolute error
Mean squared error

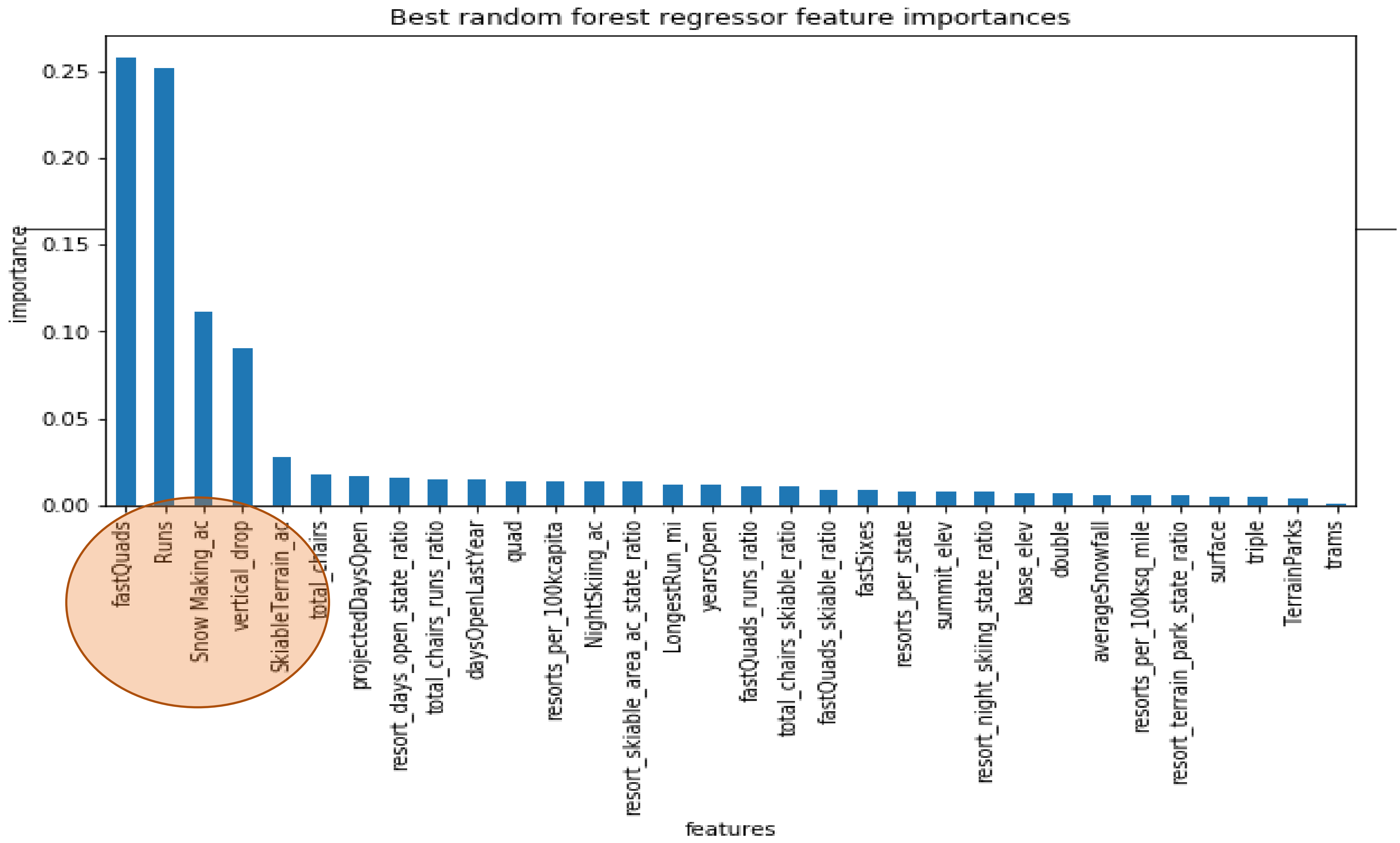
Data Pre-processing and Training
Random Forest vs. Linear Regression

MAE: 9.50



MAE: 11.79

highly **accurate** and robust	slow in generating predictions
not suffer from the overfitting problem	difficult to interpret compared with linear regression
Tolerate missing values	
Display feature importance	
Simple to implement, easy to interpret	Assume normal distribution in variables
Less complex compared with random forest	Assume the linear relationship exists
Susceptible to overfitting	Outlier can have a huge effect on the model


Best random forest regressor feature importances
importance
0.25
0.20
0.15
0.10
0.05
0.00
fastQuads
Runs
Snow Making_ac
vertical_drop
SkiableTerrain_ac
total_chairs
projectedDaysOpen
resort_days_open_state_ratio
total_chairs_runs_ratio
daysOpenLastYear
quad
resorts_per_100kcapita
NightSkiing_ac
resort_skiable_area_ac_state_ratio
LongestRun_mi
yearsOpen
fastQuads_runs_ratio
total_chairs_skiable_ratio
fastQuads_skiable_ratio
fastSixes
resorts_per_state
summit_elev
resort_night_skiing_state_ratio
base_elev
double
averageSnowfall
resorts_per_100ksq_mile
resort_terrain_park_state_ratio
surface
triple
TerrainParks
trams
features

Modeling and Deployment

I. What’s the predicted ticket price for the Big Mountain ski-resort?

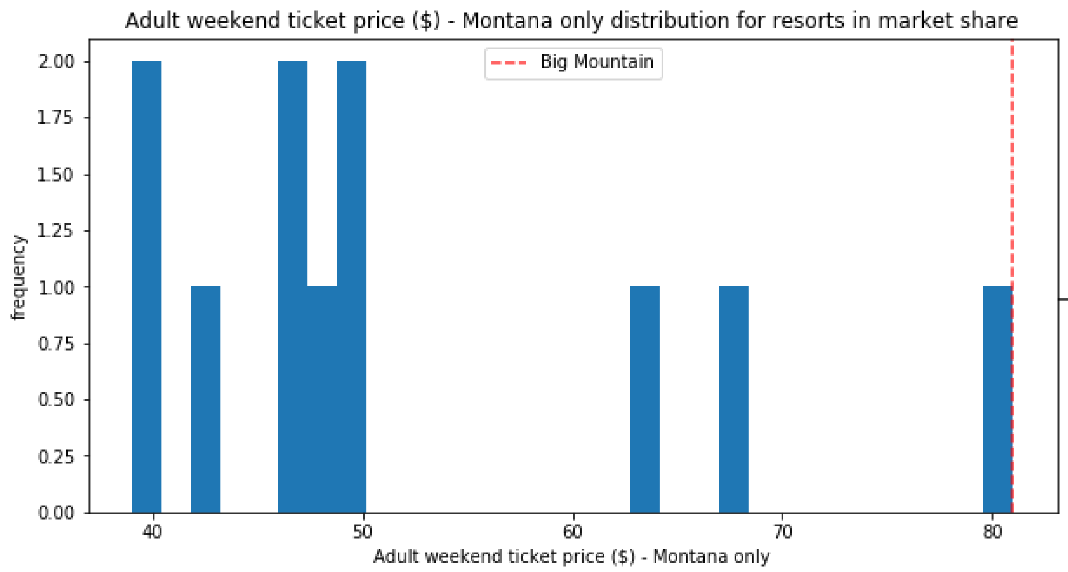
II. Which features are the most important based on the modeling results?

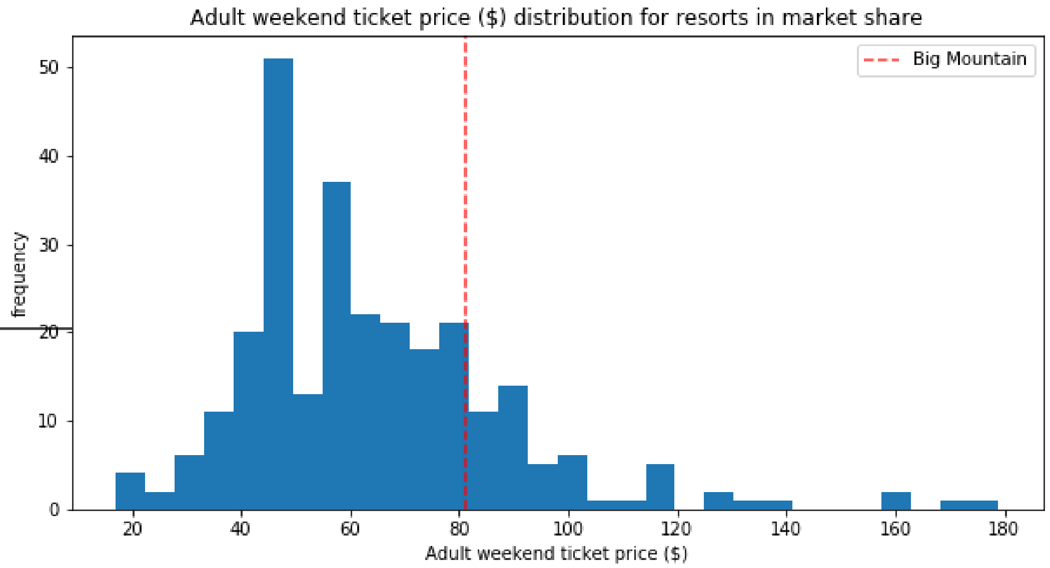
III. Does the model support other business scenarios besides increasing ticket price?

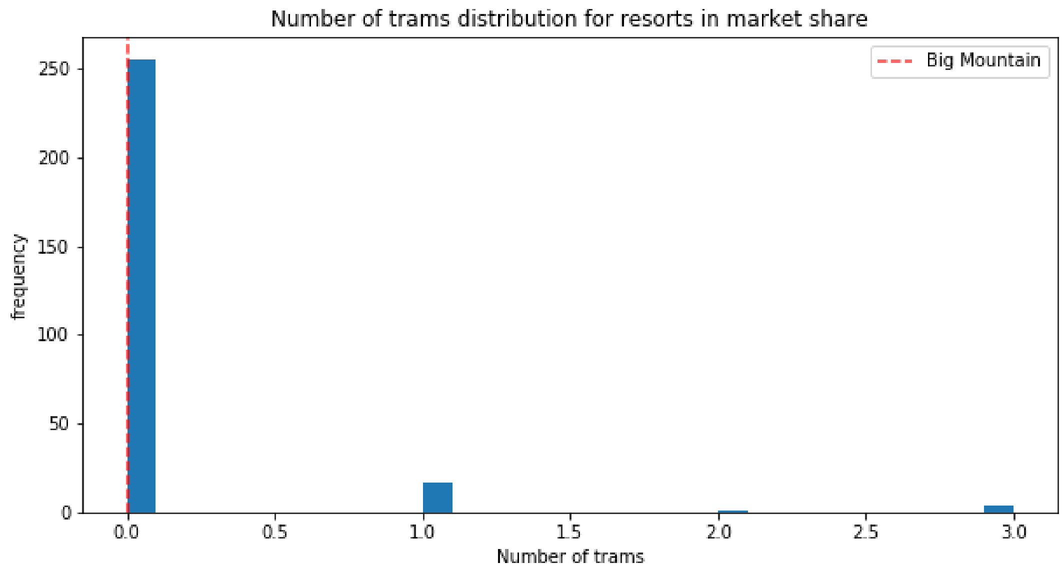
* pandas, numpy, matplotlib, seaborn, sklearn

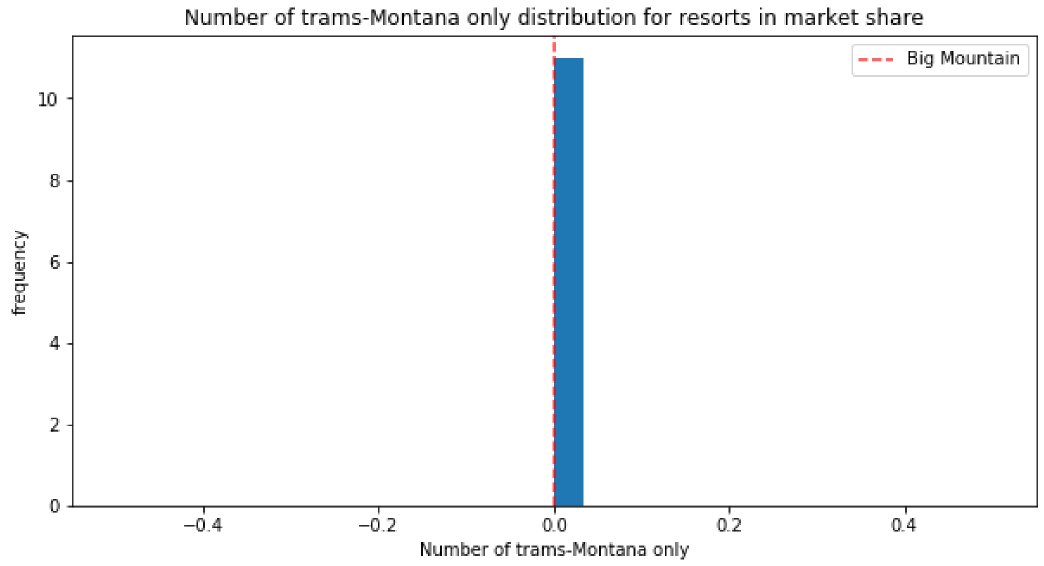
Modeling and Deployment

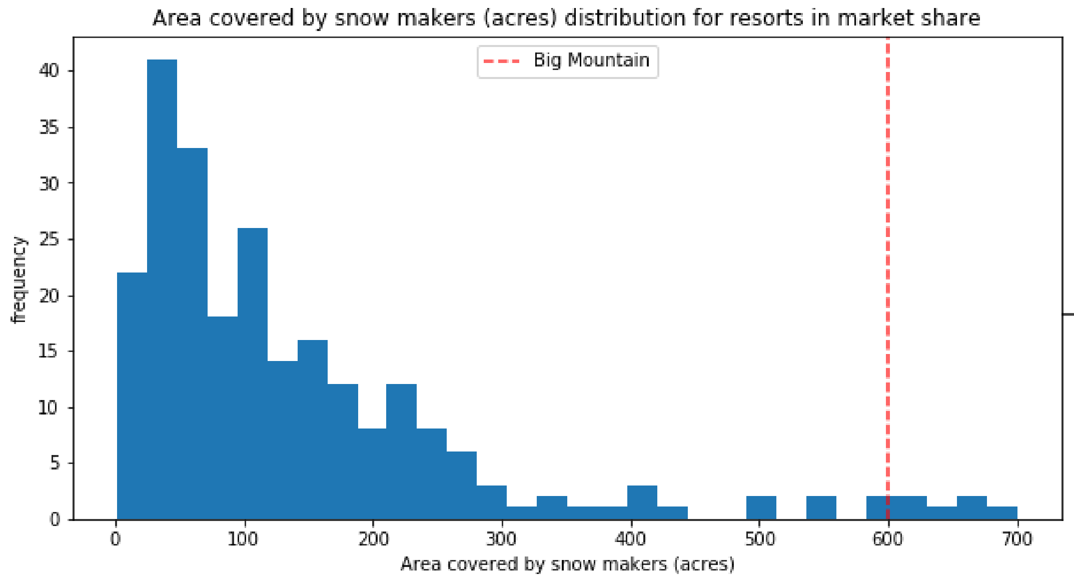
- ❑ Model prediction result is $94.22, and the actual ticket price is $81.00.
- ❑ Important features in the model: vertical_drop, SnowMaking_ac, total_chairs, fastQuads, Runs, LongestRun_mi, trams and SkiableTerrain_ac
- ❑ The Big Mountain ticket price, $80, is slightly higher than the medium of all the resorts in this study, $60, although it is the most expensive resort in Montana.
- ❑ The Big Mountain Resort is the top resort with high SnowMaking_ac, total_chairs, fastQuads, number of runs, and the most considerable amount of skiable terrain

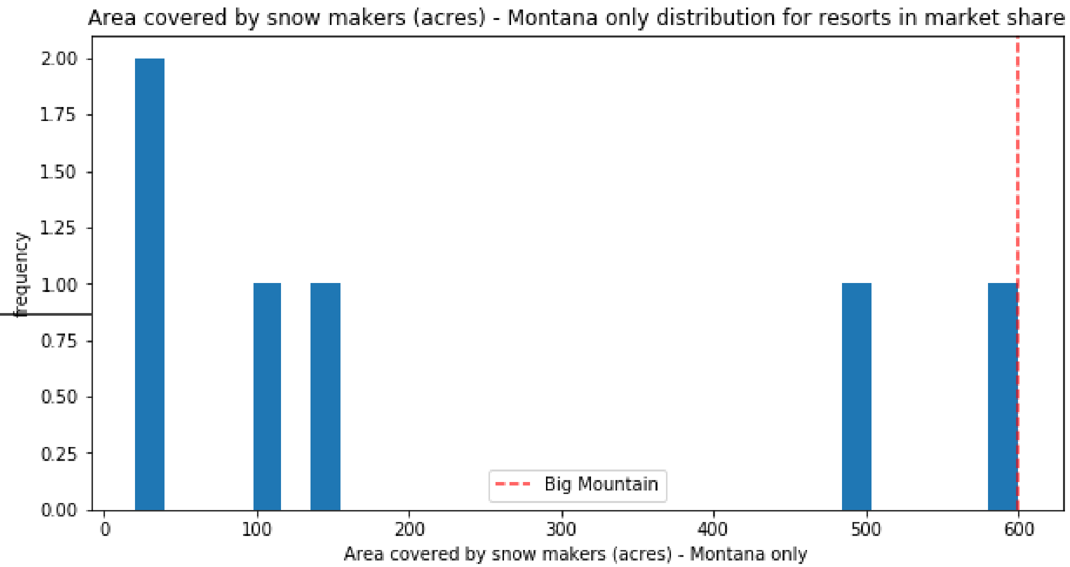

Adult weekend ticket price ($) - Montana only distribution for resorts in market share
Big Mountain
frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
40
50
60
70
80
Adult weekend ticket price ($) - Montana only

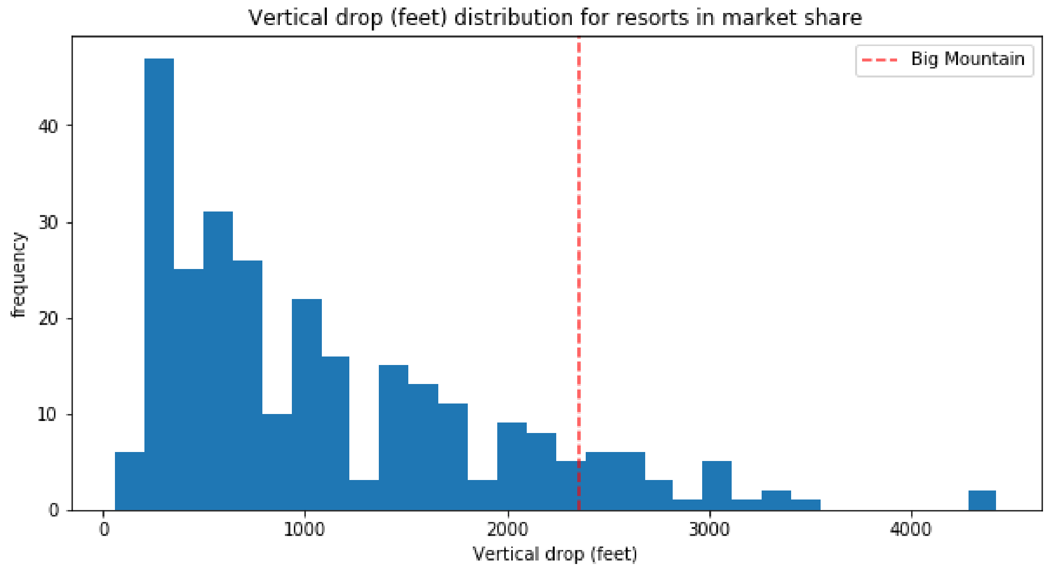

Adult weekend ticket price ($) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
50
20
40
60
80
100
120
140
160
180
Adult weekend ticket price ($)

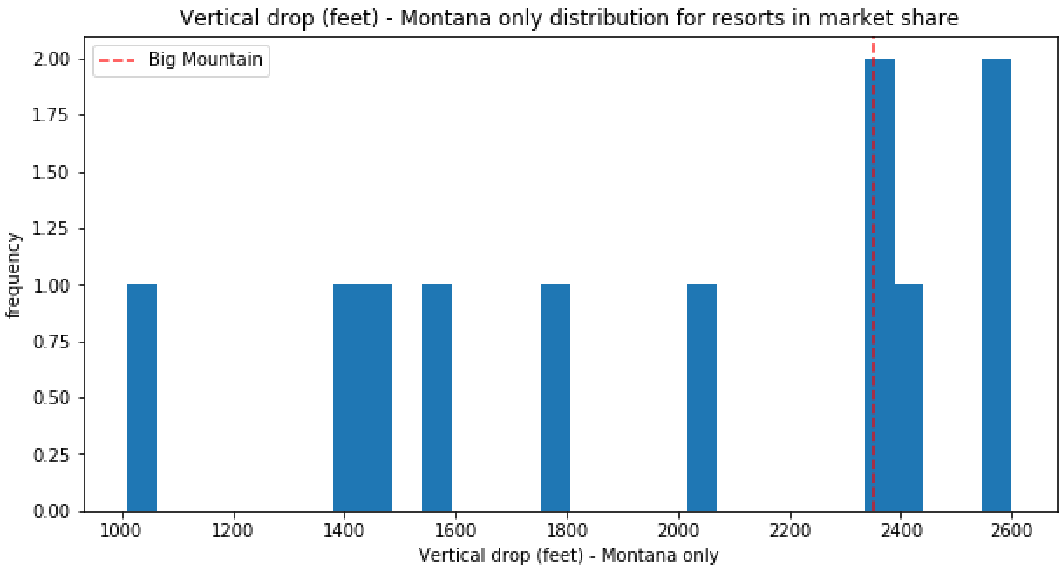

Number of trams distribution for resorts in market share
Big Mountain
frequency
0
50
100
150
200
250
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Number of trams

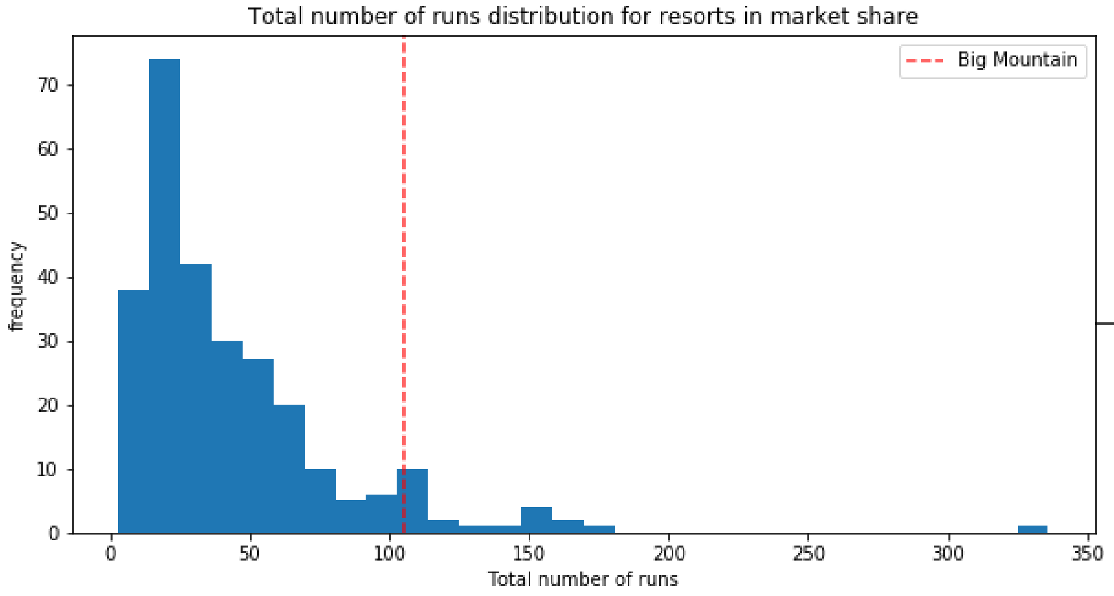

Number of trams-Montana only distribution for resorts in market share
Big Mountain
frequency
0
2
4
6
8
10
−0.4
−0.2
0.0
0.2
0.4
Number of trams-Montana only

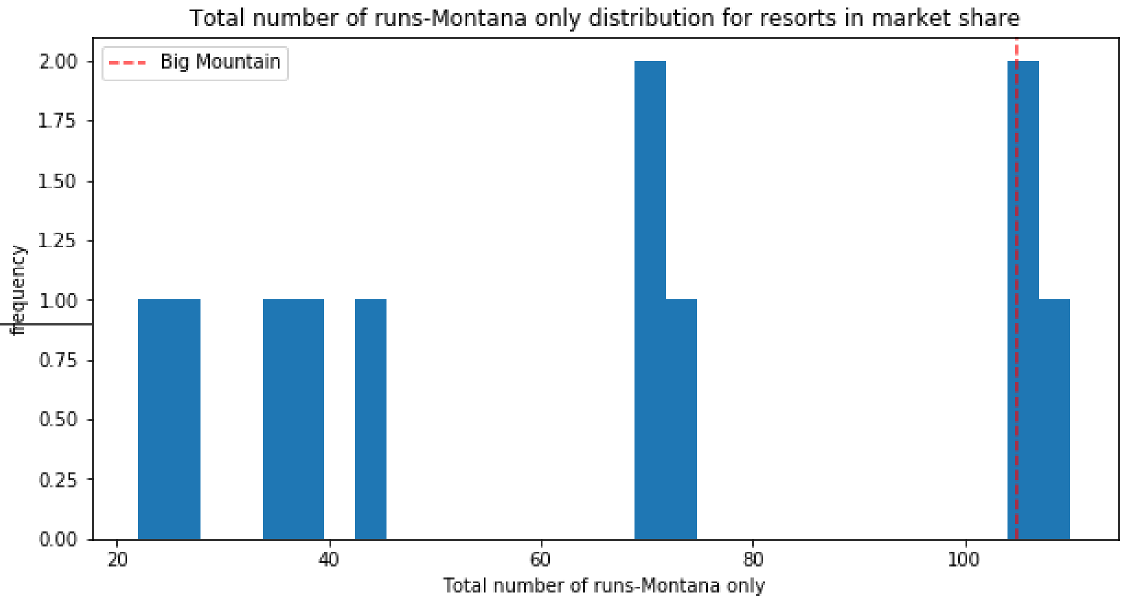

Area covered by snow makers (acres) distribution for resorts in market share
Big Mountain
frequency
Area covered by snow makers (acres)

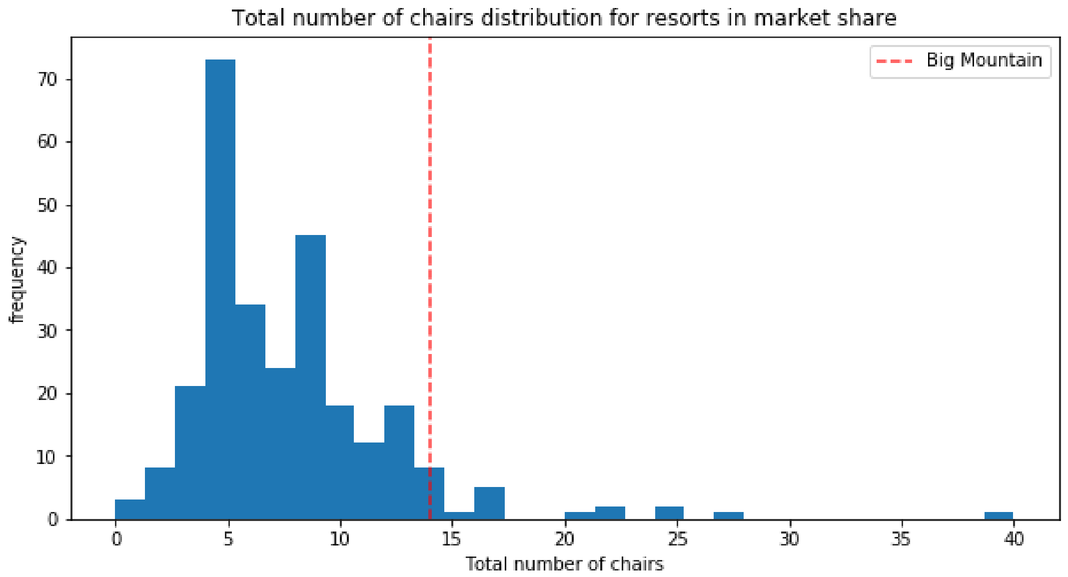

Area covered by snow makers (acres) - Montana only distribution for resorts in market share
Big Mountain
frequency
Area covered by snow makers (acres) - Montana only

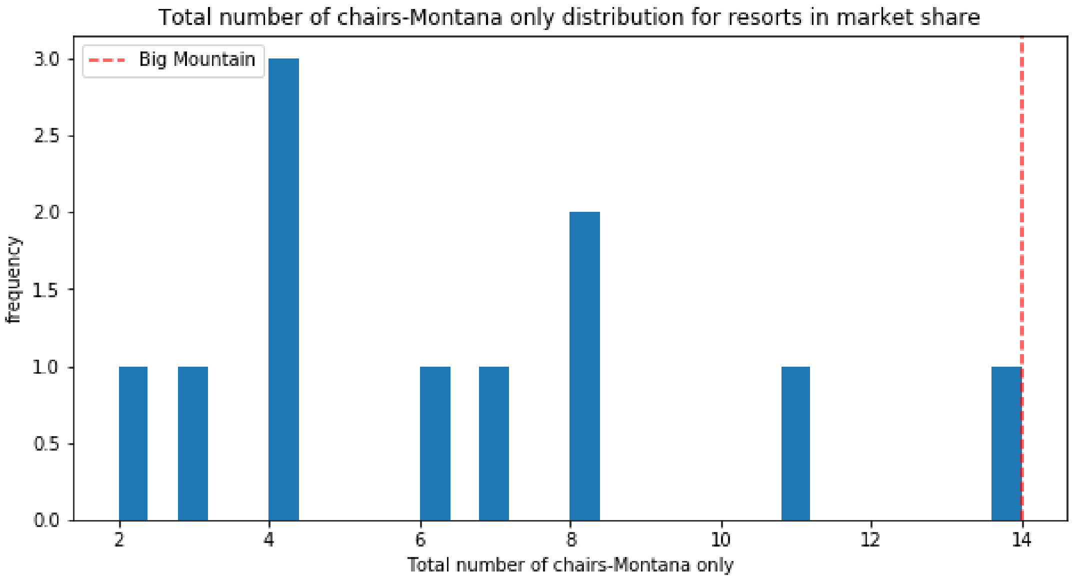

Vertical drop (feet) distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet)

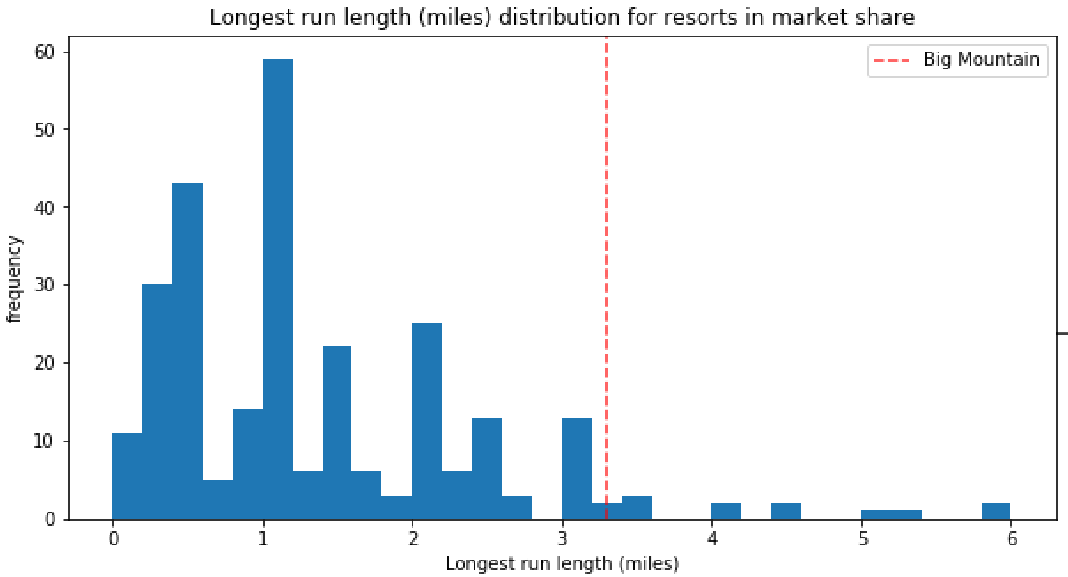

Vertical drop (feet) - Montana only distribution for resorts in market share
Big Mountain
frequency
Vertical drop (feet) - Montana only

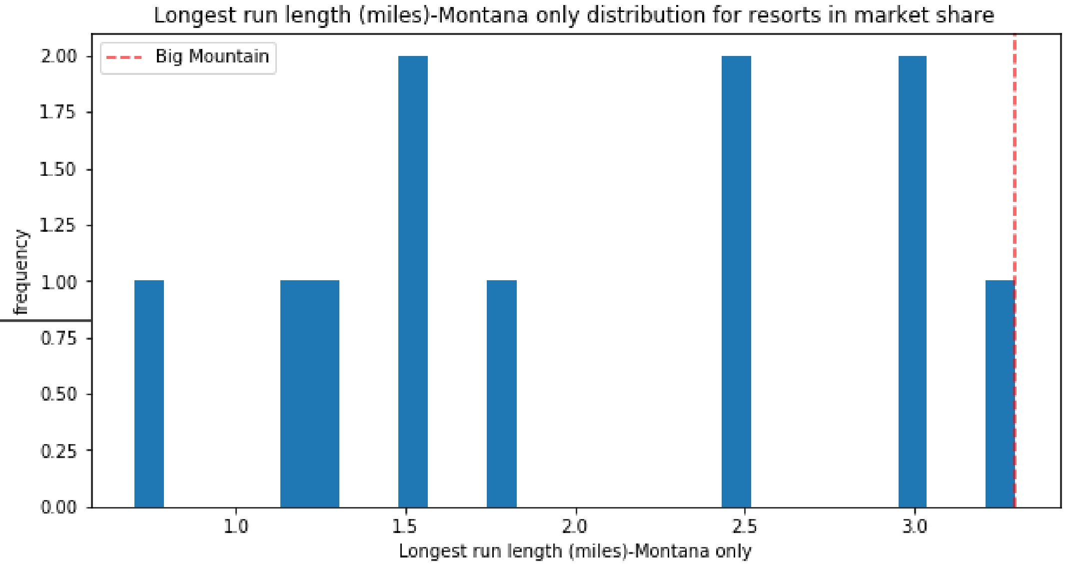

Total number of runs distribution for resorts in market share
Big Mountain
frequency
Total number of runs

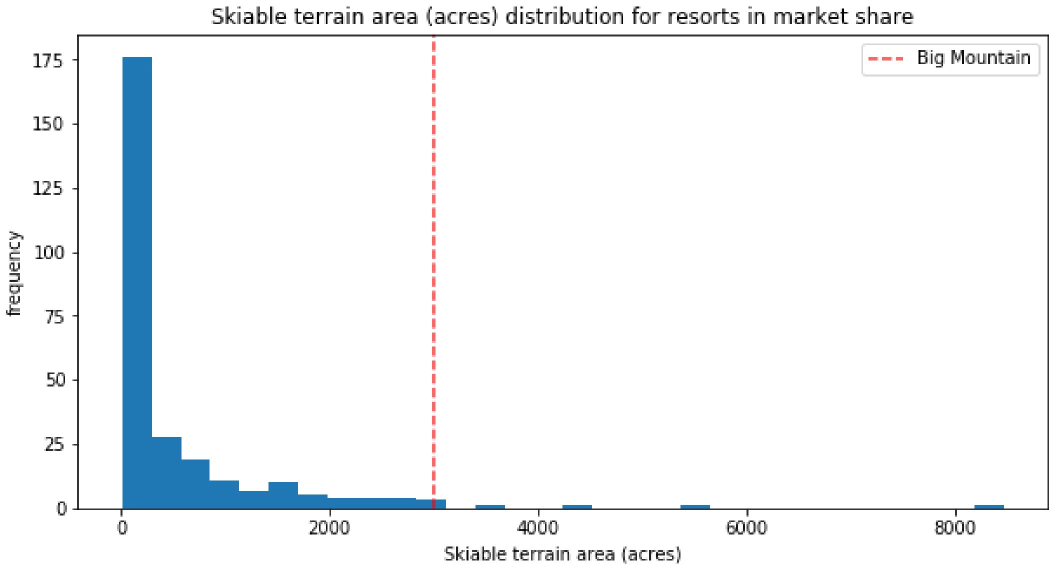

Total number of runs-Montana only distribution for resorts in market share
Big Mountain
frequency
Total number of runs-Montana only

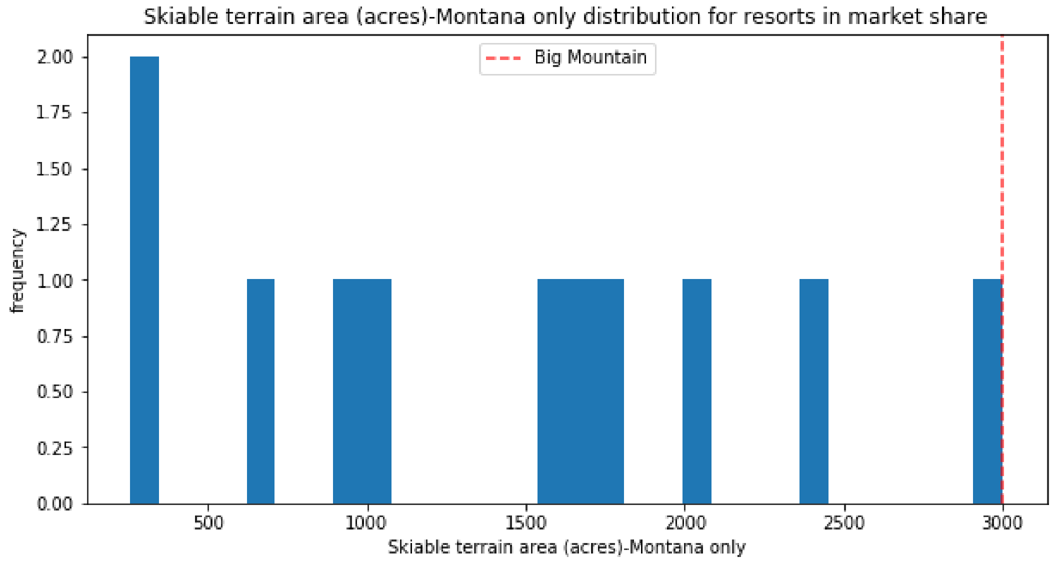

Total number of chairs distribution for resorts in market share
Big Mountain
frequency
Total number of chairs


Total number of chairs-Montana only distribution for resorts in market share
Big Mountain
frequency
Total number of chairs-Montana only


Longest run length (miles) distribution for resorts in market share
Big Mountain
frequency
0
10
20
30
40
50
60
0
1
2
3
4
5
6
Longest run length (miles)


Longest run length (miles)-Montana only distribution for resorts in market share
Big Mountain
frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
1.0
1.5
2.0
2.5
3.0
Longest run length (miles)-Montana only


Skiable terrain area (acres) distribution for resorts in market share
Big Mountain
frequency
0
25
50
75
100
125
150
175
0
2000
4000
6000
8000
Skiable terrain area (acres)


Skiable terrain area (acres)-Montana only distribution for resorts in market share
Big Mountain
frequency
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
500
1000
1500
2000
2500
3000
Skiable terrain area (acres)-Montana only

Modeling Scenarios Evaluation

I. Permanently closing down up to 10 of the least used runs.

II. Increase 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

III. Same as number 2, but adding 2 acres of snow making cover

IV. Increase the longest run by 0.2 mile to boast 3.5 miles length, requiring an additional snow making coverage of 4 acres

Key findings

S1	S2	S3	S4
Close up to 4 least used runs	Add a new run to increase the vertical drop by 150 feet , install a chair lift	Add two acres of snow making	Increase the ticket price from $81 to $94.
It will barely have effect on the revenue.	It will increase support for ticket price by $8.46. Over the season, this could be expected to amount to $14,811,594.	It will increase support for ticket price by $9.75. Over the season, this could be expected to amount to $17,068,841	It will increase the revenue by 22,750,000 dollars this season. It can completely cover the cost of the new lift.

I. Combine customer information in the study
II. Cluster analysis and market segmentation
III. Collect social media data to improve marketing accuracy

suggestion