**Q: 1** Write a summary of the exploratory data analysis above. What numerical or categorical features were in the data? Was there any pattern suggested of a relationship between state and ticket price? What did this lead us to decide regarding which features to use in subsequent modeling? What aspects of the data (e.g. relationships between features) should you remain wary of when you come to perform feature selection for modeling? Two key points that must be addressed are the choice of target feature for your modelling and how, if at all, you're going to handle the states labels in the data.

A preliminary assessments of data quality and refinement was performed and identification of the the question to be answered; how to predict adult weekend prices based on resort and state data? There were some small number of data values (resorts\_per\_100ksq\_mile and resorts\_per\_100kcapita) that gave clear choices about whether to replace values or drop a whole row. Records with missing price data were dropped, but not before making the most of the other available data to look for any patterns between the states. No discernable pattern was detected between states so the decision was to treat all states equally, therefore the state label didn’t have much relevance to answering the question.

In the data exploratory phase the objective was to identify statistical correlations between features in the data to the ticket pricing from a state summary perspective. This process required data visualizations that highlighted both the numerical and categorical features in the data. The numerical and categorical features are the following:

Categorical:

* Name
* Region
* State

Numeric:

* summit\_elev
* veritical\_drop
* base\_elev
* trams
* fastSixes
* fastQuads
* quad
* triple
* double
* surface
* total\_chairs
* Runs
* TerrainParks
* LongestRun\_mi
* SkiableTerrain\_ac
* Snow\_Making\_ac
* daysOpenLastYear
* yearsOpen
* averageSnowfall
* AdultWeekend
* projectedDaysOpen
* NightSkiing\_ac
* resorts\_per\_state
* resorts\_per\_100kcapita
* resorts\_per\_100ksq\_mile
* resort\_skiable\_area\_ac\_state\_ratio
* resort\_days\_open\_state\_ratio
* resort\_terrain\_park\_state\_ratio
* resort\_night\_skiing\_state\_ratio
* total\_chairs\_run\_ratio
* total\_chairs\_skiable\_ratio
* fastQuads\_runs\_ratio
* fastQuads\_skiable\_ratio