**Guided Capstone Project Report**

1. **Training Findings**

- The median comparison showed better results than the mean comparisons.

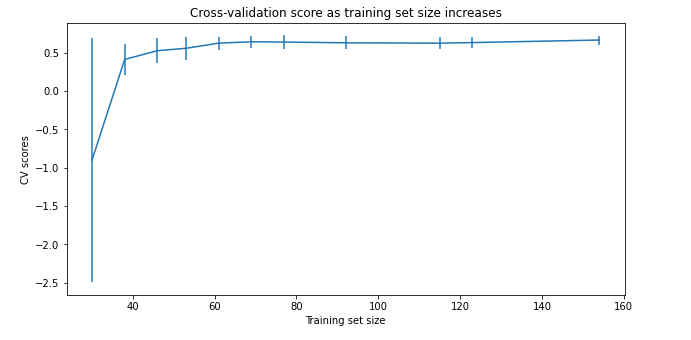
- Data was scaled using Standard Scaler and imputed with either the mean or the median.

- The Train and Test data split as 70% and 30%.

- Pipes were created for both linear regression and random regressor that showed better performance and its parameters were reset, and iterated again.

1. ***Cross Validation Results***

There's an initial rapid improvement in model scores as one would expect, but it's essentially levelled off by around a sample size of 40-50.



1. ***Conclusions***

* A comparison between the States have been explored by: total area, population and resorts per state, and other skiable data.

- The States with the highest number of resorts have the most open days status throughout the year.

- The northern and eastern States have the highest total night skiing because of shorter days in terms of hours (i.e. nights have the most hours in these regions).

- The PCA was utilized to get to the summary statistics for the state/resort differentiation, the average ticket price by state were revealed for the weekend per state.

- Rhode Island was not carried through as the NaN values were not needed.

- A correlation heat map was created to look for the correlations in all variables. The variables (summit and base elevation) were extremely correlated with the night skiing and number of resorts per capita.

- The variable “AdultWeekend” has many different correlations with other variables, such as, snow making and runs.

- Scatter plots were created to visualize the correlation of ticket prices and the other variables. The variables vertical\_drop and projected days open have the most correlations.

- the more chairs /runs and, quad lift could affect the ticket price positively.