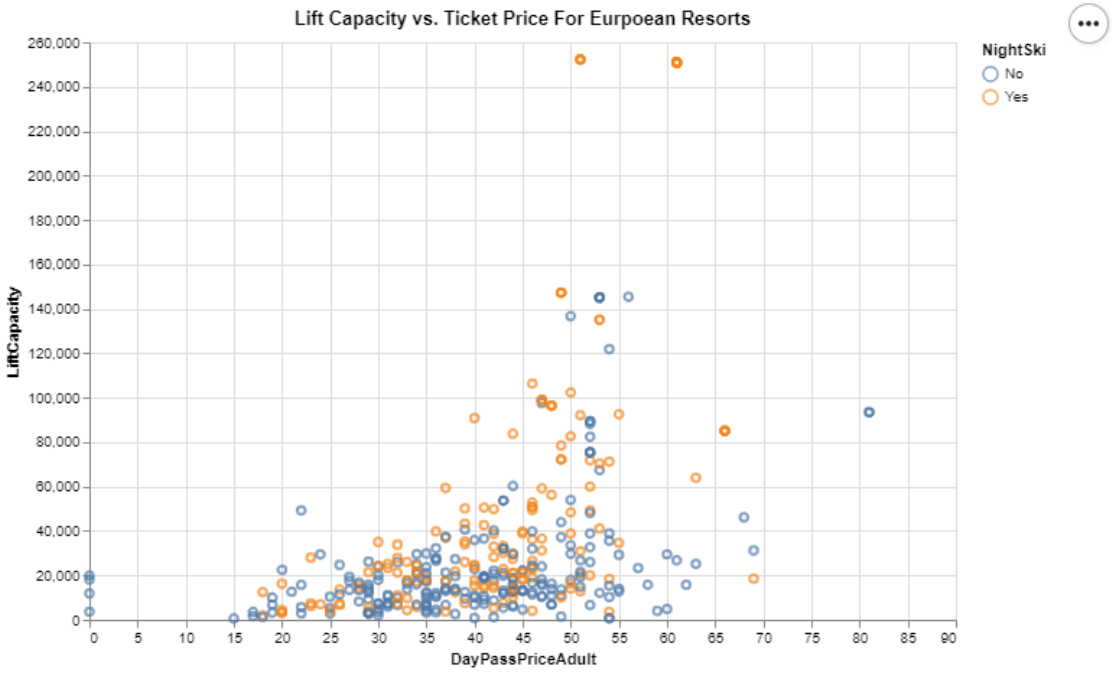
Jack Driscoll Info 4602

HW # 3 3/5/2022

For this homework assignment, I chose to work with a dataset containing information on 376 European Ski Resorts. The dataset can be found at https://www.kaggle.com/thomasnibb/european-ski-resorts, and is populated with columns including resort name, country, amount of total runs, amount of chairlifts, amount of gondolas, ticket price, and much more. The data was interesting to me as I love to ski, and am interested in the economy and functionality of the ski industry. I worked with Pandas to clean and manipulate the data, and Altair to visualize it.

For my first vis, I was interested in the price of a ticket compared to the total lift capacity of a resort. I would assume that the larger a resort, in general, the more a lift ticket would cost. I was also interested in whether there was a significant correlation between whether a resort offered night skiing or not and the cost of a ticket. I plotted the ticket price on the x, lift capacity on the y, and colored the marks by whether or not the resort offered night skiing. My visualization works in that it shows a distribution where, in general, larger resorts have higher prices and resorts that offer night skiing also have higher prices. It illustrates a trend in the data that I was attempting to convey.

```
chart = alt.Chart(resortsdf).mark_point().encode(
    x='DayPassPriceAdult',
    y='LiftCapacity',
    color = 'NightSki'
).properties(
    width = 600,
    height = 400,
    title = "Lift Capacity vs. Ticket Price For Eurpoean Resorts"
)
chart
```



In my second visualization, I plot total slope on the y and highest point on the x. I also used Altair's "transform_regression()" function to apply a best fit line to my mark point plot. I think that this visualization, although relatively simple, succeeds in conveying a clear trend. That larger (more slopes) resorts tend to have highest points that are above those of smaller (less slopes) resorts. In reflection, it perhaps may have been more effective if there was an "average altitude" column where I could look at altitude in compared to resort ski area.

```
chart = alt.Chart(resortsdf).mark_point().encode(
    x='HighestPoint',
    y='TotalSlope',
).properties( #Specifying Hieght and Width of chart
    width = 600,
    height = 400,
    title = 'Total Skiable Slope vs. Altitude'
)

chart + chart.transform_regression('HighestPoint','TotalSlope').mark_line()
```