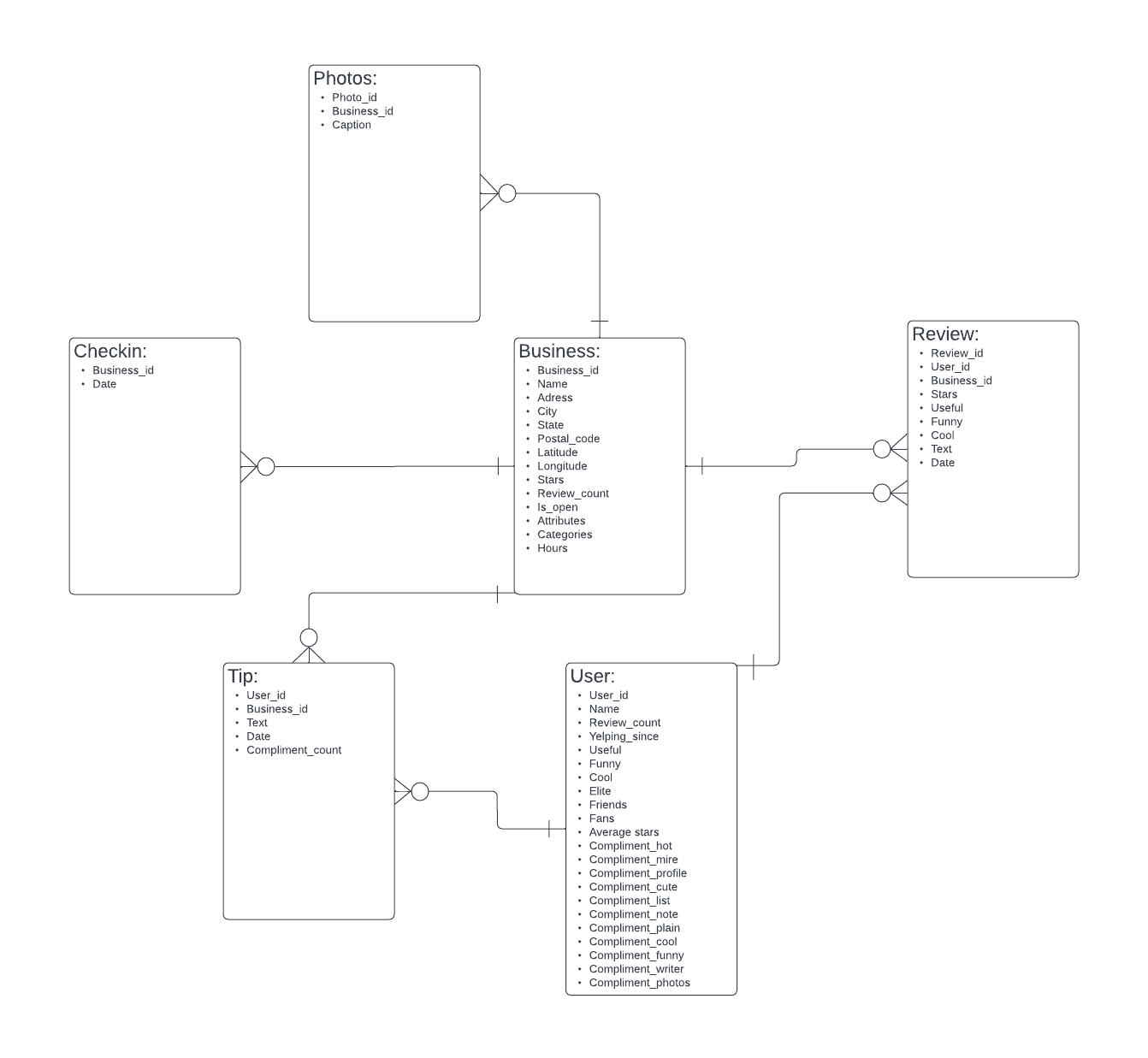
**My Dataset:** I have decided to look at the CoffeKing dataset. I am interested in how the businesses work and how we can use information about the social media users to improve business.

**ERD:**



**Hypotheses and Approach:**

1. There are couple attributes in the businesses that have a crucial impact on business star rating.

To find these attributes I will see how the star ratings different given that the attribute is on the opposite side of the spectrum (Either false/true or on a scale). I will rank them and examine.

1. People from different states have bias towards businesses that makes it harder to relate them to each other.

To find these biases I will see what are the average rating for coffee shops and for all the businesses. Then I will compare between these to find an unbiased rating and find which states like coffee the most.

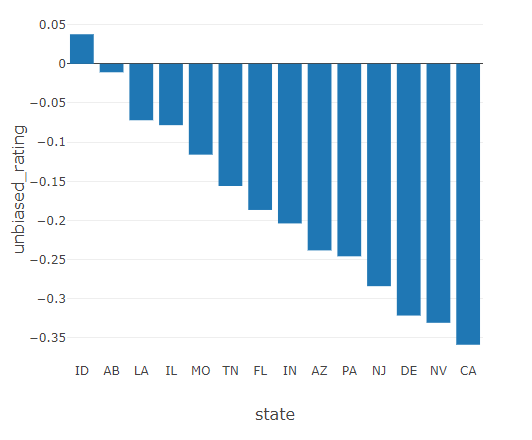
**Results:**

1. I have found that the following attributes are the most influential:

|  |  |
| --- | --- |
| Drive Thru | -1.96 star if True |
| Quiet | +1.43 star if True |
| HasTV | -0.86 star if True |
| Reservations | +0.60 stars if True |
| Outdoor Seating | +0.52 stars if True |

1. The states that value coffee the most are:

|  |  |  |  |
| --- | --- | --- | --- |
| state | coffee\_rating | business\_rating | unbiased\_rating |
| ID | 3.771084 | 3.73328 | 0.037804 |
| AB | 3.451493 | 3.462264 | -0.01077 |
| LA | 3.626506 | 3.698541 | -0.07203 |
| IL | 3.291667 | 3.369972 | -0.07831 |
| MO | 3.451299 | 3.567261 | -0.11596 |
| TN | 3.422438 | 3.578438 | -0.156 |
| FL | 3.439863 | 3.626462 | -0.1866 |
| IN | 3.39911 | 3.602895 | -0.20379 |
| AZ | 3.380597 | 3.618833 | -0.23824 |
| PA | 3.357783 | 3.603579 | -0.2458 |
| NJ | 3.186981 | 3.470915 | -0.28393 |
| DE | 3.044444 | 3.365892 | -0.32145 |
| NV | 3.439086 | 3.769794 | -0.33071 |
| CA | 3.717391 | 4.076138 | -0.35875 |



**Hypotheses Discussion:**

1. My initial hypothesis was unspecified as I needed to examine the data in detail in order to answer my very broad question. However, I specified it to its current state in Milestone 2.
2. Hypothesis 2 evolved a lot during the process. It was unspecified and based on solely the data of coffee industry. However, as the discovery of data progressed, I have decided to look at the rating of coffeeshops related to the average rating of all the business in the state as the simple comparison of states would be unreliable due to customer bias.

**Metrics Discussion:**

1. I realized that I cannot examine the attributes using the AB test as there were no trails within the data. Thus, I decided to measure the importance of the attributes using simple calculation and relation to each other.
2. Metric for this thesis evolved from comparison of ratings between states to comparison of relations of average ratings of all businesses to coffeeshops between states.

**Relationships Discussion:**

It was discovered that the attributes are related to average ratings and there are attributes that are key to achieving higher ratings. Another key finding is that there is customer bias based on the location of the customers. This led to determining which states are most coffee business liking without bias.

**Recommendations and Actions:**

There are two key finding in this study. First one is that when building coffee shop there are key attributes to which the business has to put special attention. The coffeeshops should be built without TVs, they should be as quiet as possible, should allow for reservations, and should have outdoor seating. DriveThru’s are usually the weakest link of a coffeeshop, so if they are built, they should be taken care of with special attention.

The second finding is that if the new location of a coffeeshop is built, it will be most appreciated in Idaho as it has the highest unbiased rating. What is interesting Even though Californian coffeeshops are second highest rated on yelp, they have on average the lowest rating compared to other businesses in CA.