

## Use data, primarily from `sliced_data.csv` to predict the `review_overall` column for a held out brewery and its beers

We've held out reviews of Surly Brewing Company's beers from a select group of reviewers. The object is to predict what these reviewers rated Surly Brewing Company's beers based on how they reviewed other beers.

Please be mindful of the other held out features.

### `sliced_holdout_data.csv`

All columns in `sliced_data.csv` are the same as in this file **EXCEPT FOR:**

```
review_overall
review_aroma
review_appearance
review_palate
review_taste
```

**THE ABOVE FEATURES ARE NOT IN THE HOLDOUT SET**

### `sliced_data.csv`

```
brewery_id : int unique brewery ID
brewery_city : str brewery city
brewery_state : str brewery state (if US)
brewery_country : str brewery country
brewery_name : str brewery name
review_time : int unix timestamp of review review_overall : float overall rating (what
was the overall experience of the beer?)
review_aroma : float the smell (malt, hops, yeast, other aromatics)
review_appearance : float the look (beer color, clarity, head retention, lacing on the
glass)
review_profilename : str reviewer username
beer_style : str specific type of beer (i.e. American Pale Ale, Berlinerweiss)
review_palate : float mouthfeel (body, carbonation, warmth, creaminess, astringency,
other palate sensations)
review_taste : float flavor of malt, hops, fermentation byproducts, balance, finish,
aftertaste, other flavor characteristics
beer_name : str name of the beer
beer_abv : float percent of alcohol by volume
beer_beerid : unique beer ID
beer_category : str broad type of beer (i.e. Pale Ale, Sour)
beer_availability : str whether a beer is available year-round or otherwise
brewery_review_time_mean : useless aggregation of unix timestamp ints
brewery_review_overall_mean : aggregated overall rating of all beers at a brewery
brewery_review_aroma_mean : aggregated aroma rating of all beers at a brewery
brewery_review_appearance_mean : aggregated appearance rating of all beers at a
brewery
brewery_review_palate_mean : aggregated palate rating of all beers at a brewery
brewery_review_taste_mean : aggregated taste rating of all beers at a brewery
```

`brewery_beer_abv_mean` : aggregated abv of all beers at a brewery  
`beer_category_review_time_mean` : useless aggregate  
`beer_category_review_overall_mean` : aggregated overall rating of all beers in a given beer category  
`beer_category_review_aroma_mean` : aggregated aroma rating of all beers in a given beer category  
`beer_category_review_appearance_mean` : aggregated appearance rating of all beers in a given beer category  
`beer_category_review_palate_mean` : aggregated palate rating of all beers in a given beer category  
`beer_category_review_taste_mean` : aggregated taste rating of all beers in a given beer category  
`beer_category_beer_abv_mean` : aggregated abv of all beers in a given beer category

## **reviewer\_info.csv**

More aggregated info about the reviewers can be found in this dataset. There are more reviewers in this dataset than there are in the sliced dataset. The data aggregates over reviewers and their ratings for beer categories. The idea is a given reviewer may prefer certain categories of beer more than others.

`review_profilename` : str reviewer username  
`beer_category` : str the category of beer reviewed the given user  
`reviewer_number_of_reviews` : int number of reviews for a given beer category by a specific reviewer  
`reviewer_review_time_mean` : useless unix timestamp  
`reviewer_review_overall_mean` : mean aggregated overall rating of a beer category for a specified reviewer  
`reviewer_review_aroma_mean` : mean aggregated aroma rating of a beer category for a specified reviewer  
`reviewer_review_appearance_mean` : mean aggregated appearance rating of a beer category for a specified reviewer  
`reviewer_review_palate_mean` : mean aggregated palate rating of a beer category for a specified reviewer  
`reviewer_review_taste_mean` : mean aggregated taste rating of a beer category for a specified reviewer  
`reviewer_beer_abv_mean` : mean aggregated abv of a beer category for a specified reviewer  
`reviewer_review_time_median` : `reviewer_review_overall_median` : median (50%tile) aggregated overall rating of a beer category for a specified reviewer  
`reviewer_review_aroma_median` : median (50%tile) aggregated aroma rating of a beer category for a specified reviewer  
`reviewer_review_appearance_median` : median (50%tile) aggregated appearance rating of a beer category for a specified reviewer  
`reviewer_review_palate_median` : median (50%tile) aggregated palate rating of a beer category for a specified reviewer  
`reviewer_review_taste_median` : median (50%tile) aggregated taste rating of a beer category for a specified reviewer  
`reviewer_beer_abv_median` : median (50%tile) aggregated abv of a beer category for a specified reviewer  
`reviewer_review_time_sd` : `reviewer_review_overall_sd` : standard deviation aggregated overall rating of a beer category for a specified reviewer  
`reviewer_review_aroma_sd` : standard deviation aggregated aroma rating of a beer

category for a specified reviewer

reviewer\_review\_appearance\_sd : standard deviation aggregated appearance rating of a beer category for a specified reviewer

reviewer\_review\_palate\_sd : standard deviation aggregated palate rating of a beer category for a specified reviewer

reviewer\_review\_taste\_sd : standard deviation aggregated taste rating of a beer category for a specified reviewer

reviewer\_beer\_abv\_sd : standard deviation aggregated abv of a beer category for a specified reviewer

reviewer\_review\_time\_05p : reviewer\_review\_overall\_05p : 5%tile aggregated overall rating of a beer category for a specified reviewer

reviewer\_review\_aroma\_05p : 5%tile aggregated aroma rating of a beer category for a specified reviewer

reviewer\_review\_appearance\_05p : 5%tile aggregated appearance rating of a beer category for a specified reviewer

reviewer\_review\_palate\_05p : 5%tile aggregated palate rating of a beer category for a specified reviewer

reviewer\_review\_taste\_05p : 5%tile aggregated taste rating of a beer category for a specified reviewer

reviewer\_beer\_abv\_05p : 5%tile aggregated abv of a beer category for a specified reviewer

reviewer\_review\_time\_25p : reviewer\_review\_overall\_25p : 25%tile aggregated overall rating of a beer category for a specified reviewer

reviewer\_review\_aroma\_25p : 25%tile aggregated aroma rating of a beer category for a specified reviewer

reviewer\_review\_appearance\_25p : 25%tile aggregated appearance rating of a beer category for a specified reviewer

reviewer\_review\_palate\_25p : 25%tile aggregated palate rating of a beer category for a specified reviewer

reviewer\_review\_taste\_25p : 25%tile aggregated taste rating of a beer category for a specified reviewer

reviewer\_beer\_abv\_25p : 25%tile aggregated abv of a beer category for a specified reviewer

reviewer\_review\_time\_75p : reviewer\_review\_overall\_75p : 75%tile aggregated overall rating of a beer category for a specified reviewer

reviewer\_review\_aroma\_75p : 75%tile aggregated aroma rating of a beer category for a specified reviewer

reviewer\_review\_appearance\_75p : 75%tile aggregated appearance rating of a beer category for a specified reviewer

reviewer\_review\_palate\_75p : 75%tile aggregated palate rating of a beer category for a specified reviewer

reviewer\_review\_taste\_75p : 75%tile aggregated taste rating of a beer category for a specified reviewer

reviewer\_beer\_abv\_75p : 75%tile aggregated abv of a beer category for a specified reviewer

reviewer\_review\_time\_95p : reviewer\_review\_overall\_95p : 95%tile aggregated overall rating of a beer category for a specified reviewer

reviewer\_review\_aroma\_95p : 95%tile aggregated aroma rating of a beer category for a specified reviewer

reviewer\_review\_appearance\_95p : 95%tile aggregated appearance rating of a beer category for a specified reviewer

reviewer\_review\_palate\_95p : 95%tile aggregated palate rating of a beer category for a

specified reviewer

`reviewer_review_taste_95p` : 95%tile aggregated taste rating of a beer category for a specified reviewer

`reviewer_beer_abv_95p` : 95%tile aggregated abv of a beer category for a specified reviewer

`reviewer_review_time_min` : `reviewer_review_overall_min` : minimum aggregated overall rating of a beer category for a specified reviewer

`reviewer_review_aroma_min` : minimum aggregated aroma rating of a beer category for a specified reviewer

`reviewer_review_appearance_min` : minimum aggregated appearance rating of a beer category for a specified reviewer

`reviewer_review_palate_min` : minimum aggregated palate rating of a beer category for a specified reviewer

`reviewer_review_taste_min` : minimum aggregated taste rating of a beer category for a specified reviewer

`reviewer_beer_abv_min` : minimum aggregated abv of a beer category for a specified reviewer

`reviewer_review_time_max` : `reviewer_review_overall_max` : maximum aggregated overall rating of a beer category for a specified reviewer

`reviewer_review_aroma_max` : maximum aggregated aroma rating of a beer category for a specified reviewer

`reviewer_review_appearance_max` : maximum aggregated appearance rating of a beer category for a specified reviewer

`reviewer_review_palate_max` : maximum aggregated palate rating of a beer category for a specified reviewer

`reviewer_review_taste_max` : maximum aggregated taste rating of a beer category for a specified reviewer

`reviewer_beer_abv_max` : maximum aggregated abv of a beer category for a specified reviewer

## **brewery\_and\_beer\_info.csv**

More aggregated info on beers from a brewery in this dataset. Has more beers and breweries than the sliced dataset -- does include the target brewery! The idea is that a specific brewery may craft certain beer categories well or not well.

`brewery_id` : int unique brewery ID

`beer_category` : str broad beer category

`brewery_beer_category_number_of_reviews` : int number of reviews for a given beer category at a specific brewery

`brewery_beer_category_review_time_mean` : useless unix timestamp

`brewery_beer_category_review_overall_mean` : mean aggregated overall rating of a beer category for a specified brewery

`brewery_beer_category_review_aroma_mean` : mean aggregated aroma rating of a beer category for a specified brewery

`brewery_beer_category_review_appearance_mean` : mean aggregated appearance rating of a beer category for a specified brewery

`brewery_beer_category_review_palate_mean` : mean aggregated palate rating of a beer category for a specified brewery

`brewery_beer_category_review_taste_mean` : mean aggregated taste rating of a beer category for a specified brewery

`brewery_beer_category_beer_abv_mean` : mean aggregated abv of a beer category for a specified brewery

brewery\_beer\_category\_review\_time\_median :  
brewery\_beer\_category\_review\_overall\_median : median aggregated overall rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_aroma\_median : median aggregated aroma rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_appearance\_median : median aggregated appearance rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_palate\_median : median aggregated palate rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_taste\_median : median aggregated taste rating of a beer category for a specified brewery  
brewery\_beer\_category\_beer\_abv\_median : median aggregated abv of a beer category for a specified brewery  
brewery\_beer\_category\_review\_time\_sd : brewery\_beer\_category\_review\_overall\_sd : standard deviation aggregated overall rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_aroma\_sd : standard deviation aggregated aroma rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_appearance\_sd : standard deviation aggregated appearance rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_palate\_sd : standard deviation aggregated palate rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_taste\_sd : standard deviation aggregated taste rating of a beer category for a specified brewery  
brewery\_beer\_category\_beer\_abv\_sd : standard deviation aggregated abv of a beer category for a specified brewery  
brewery\_beer\_category\_review\_time\_05p : brewery\_beer\_category\_review\_overall\_05p : 5%tile aggregated overall rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_aroma\_05p : 5%tile aggregated aroma rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_appearance\_05p : 5%tile aggregated appearance rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_palate\_05p : 5%tile aggregated palate rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_taste\_05p : 5%tile aggregated taste rating of a beer category for a specified brewery  
brewery\_beer\_category\_beer\_abv\_05p : 5%tile aggregated abv of a beer category for a specified brewery  
brewery\_beer\_category\_review\_time\_25p : brewery\_beer\_category\_review\_overall\_25p : 25%tile aggregated overall rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_aroma\_25p : 25%tile aggregated aroma rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_appearance\_25p : 25%tile aggregated appearance rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_palate\_25p : 25%tile aggregated palate rating of a beer category for a specified brewery  
brewery\_beer\_category\_review\_taste\_25p : 25%tile aggregated taste rating of a beer category for a specified brewery  
brewery\_beer\_category\_beer\_abv\_25p : 25%tile aggregated abv of a beer category for a specified brewery  
brewery\_beer\_category\_review\_time\_75p : brewery\_beer\_category\_review\_overall\_75p : 75%tile aggregated overall rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_aroma\_75p : 75%tile aggregated aroma rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_appearance\_75p : 75%tile aggregated appearance rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_palate\_75p : 75%tile aggregated palate rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_taste\_75p : 75%tile aggregated taste rating of a beer category for a specified brewery

brewery\_beer\_category\_beer\_abv\_75p : 75%tile aggregated abv of a beer category for a specified brewery

brewery\_beer\_category\_review\_time\_95p : brewery\_beer\_category\_review\_overall\_95p : 95%tile aggregated overall rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_aroma\_95p : 95%tile aggregated aroma rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_appearance\_95p : 95%tile aggregated appearance rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_palate\_95p : 95%tile aggregated palate rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_taste\_95p : 95%tile aggregated taste rating of a beer category for a specified brewery

brewery\_beer\_category\_beer\_abv\_95p : 95%tile aggregated abv of a beer category for a specified brewery

brewery\_beer\_category\_review\_time\_min : brewery\_beer\_category\_review\_overall\_min : minimum aggregated overall rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_aroma\_min : minimum aggregated aroma rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_appearance\_min : minimum aggregated appearance rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_palate\_min : minimum aggregated palate rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_taste\_min : minimum aggregated taste rating of a beer category for a specified brewery

brewery\_beer\_category\_beer\_abv\_min : minimum aggregated abv of a beer category for a specified brewery

brewery\_beer\_category\_review\_time\_max : brewery\_beer\_category\_review\_overall\_max : maximum aggregated overall rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_aroma\_max : maximum aggregated aroma rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_appearance\_max : maximum aggregated appearance rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_palate\_max : maximum aggregated palate rating of a beer category for a specified brewery

brewery\_beer\_category\_review\_taste\_max : maximum aggregated taste rating of a beer category for a specified brewery

brewery\_beer\_category\_beer\_abv\_max : maximum aggregated abv of a beer category for a specified brewery