

# How much for the 'Dank Nuggs'?

AJ Davis Project Luther



# Problem

- Can we predict cannabis flower prices at dispensaries?
- Why relevant?
  - Are dispensaries pricing their flower products optimally?



|              |        |       |     |     |     |
|--------------|--------|-------|-----|-----|-----|
| Golden Haze  | 56.01% | 0.27% | 100 | 100 | 100 |
| Grand Hindu  | 59.01% | 0.24% | 100 | 100 | 100 |
| Haze         | 55.90% | 0.38% | 100 | 100 | 100 |
| Hindu Kush   | 59.98% | 0.93% | 100 | 100 | 100 |
| Master Kush  | 62.45% | 1.79% | 100 | 100 | 100 |
| Purple Kush  | 49.56% | 0.95% | 100 | 100 | 100 |
| Purple Wreck | 54.03% | 0.48% | 100 | 100 | 100 |
| Red Dragon   | 53.16% | 1.35% | 100 | 100 | 100 |
| Shaman       | 42.13% | 0.08% | 100 | 100 | 100 |
|              | 51.54% | 1.61% | 100 | 100 | 100 |
|              | 49.29% | 0.24% | 100 | 100 | 100 |

# Approach

- Data
- Supervised Machine Learning

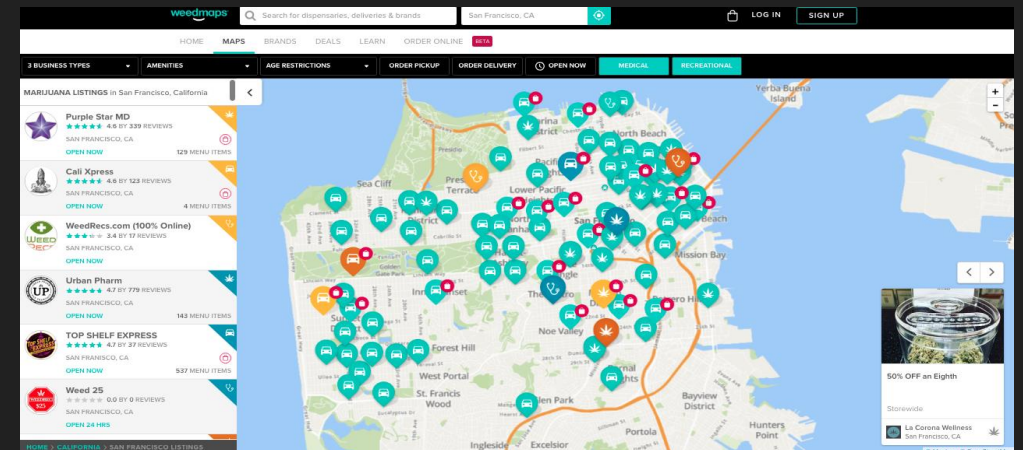
The screenshot displays the website for Vallejo Holistic Health Center (VHHC). The header includes navigation links: HOME, MAPS, BRANDS, DEALS, LEARN, ORDER ONLINE, and a GET IT button. The main content area features the center's logo, a 4.3-star rating, phone number (707) 652-5018, and operating hours (9:00am - 8:00pm). A map shows the location in Vallejo, CA. Below this, a description states: "We have a great selection of CBD products and some of the best Top Shelf items in the Bay Area! All Veterans". A "View Map, Hours, and Other Details" link is provided. The "MENU" tab is selected, showing a search bar and a "Sort by..." dropdown. The menu is categorized by "Indica" and lists three products:

| Product Category | Brands     | Price Range | Verified Products        |
|------------------|------------|-------------|--------------------------|
| All Products     | All Brands | Any         | <input type="checkbox"/> |

| Product Name                    | Price |
|---------------------------------|-------|
| Indica   Flow Kana Jah Goo      | \$28  |
| Indica   Flow Kana Pineapple OG | \$28  |
| Indica   Flow Kana Skywalker OG | \$70  |

# Pipeline

- Scrape the data
  - Cities → Geocodes → Dispensary Names → Dispensary Menus
- Identify apriori model specifications
- Check specification scores for various models
- Fit to test data and check for overfitting



# Feature Space

- THC Content
- Cross-section of dispensaries
- Dispensary ratings
- Dispensary license type
- Strain name
- Strain type
- Rare Strains

# Final Model Specifications

Specification 1

$$\boxed{Price} = THC \% + Rating_D + License_D + \sum_{r=1}^R D + \sum_{t=1}^T D$$

Specification 2

$$\boxed{\log}(Price) = THC \% + Rating_D + License_D + \sum_{r=1}^R D + \sum_{t=1}^T D$$

Specification 3

$$Price = THC \% + THC \%^2 + Rating_D + License_D + \sum_{r=1}^R D + \sum_{t=1}^T D$$

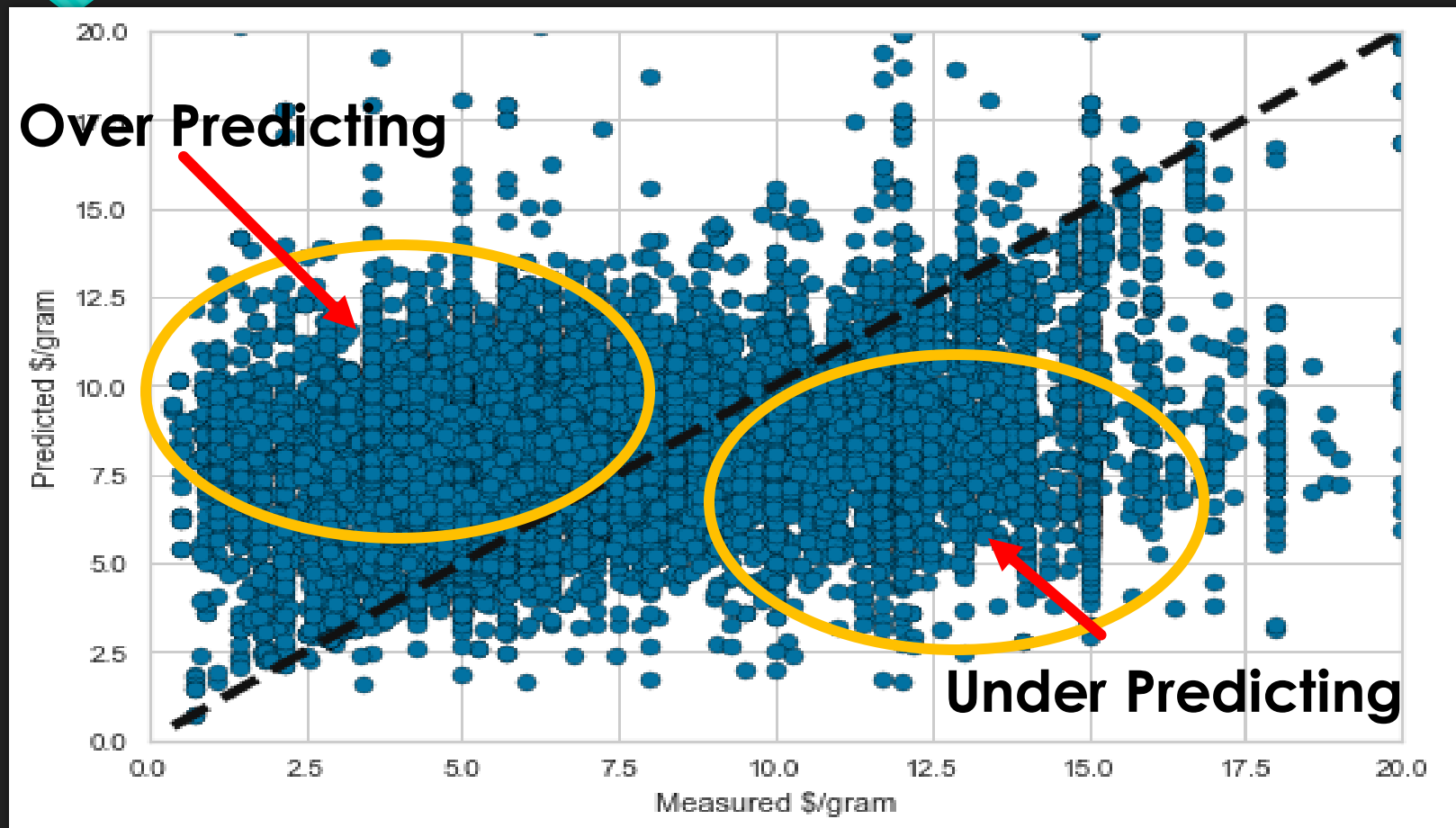
# Results

| Model (Training)                     | Specification 1<br>Score (R2) | Specification 2<br>Score (R2) | Specification 3<br>Score (R2) |
|--------------------------------------|-------------------------------|-------------------------------|-------------------------------|
| Linear Regression                    | 0.10                          | 0.13                          | 0.22                          |
| Ridge Regression                     | 0.09                          | -1.44                         | -8.51                         |
| Lasso Regression                     | 0.05                          | 0.00                          | -0.19                         |
| Robust Regression                    | -0.14                         | -1.34                         | -0.86                         |
| Epsilon Insensitive Regression       | -1.18                         | -1.42                         | -2.13                         |
| CART                                 | 0.30                          | 0.21                          | 0.34                          |
| Extra Trees                          | 0.25                          | 0.14                          | 0.21                          |
| Random Forest Regression             | 0.41                          | 0.30                          | 0.48                          |
| ADA Boosted Trees Regression         | -0.08                         | 0.2                           | 0.16                          |
| Gradient Boosted Trees<br>Regression | 0.39                          | 0.27                          | 0.40                          |

Test  
Score

0.41

# How well is the best model predicting?





# Discussion and Next Steps

- Critical omitted features
  - "Nose"
  - Density
  - CBD?
  - Quantity
  - Demand
- Better string processing for strain data
- Look at other dispensary products
- Hyperparameter tuning

