

Eileen's Coffeeshop



Project Best Buds



The Data Set

	Strain	Type	Rating	Effects	Flavor	Description
0	100-Og	hybrid	4.0	Creative,Energetic,Tingly,Euphoric,Relaxed	Earthy,Sweet,Citrus	\$100 OG is a 50/50 hybrid strain that packs a ...
1	98-White-Widow	hybrid	4.7	Relaxed,Aroused,Creative,Happy,Energetic	Flowery,Violet,Diesel	The '98 Aloha White Widow is an especially pot...
2	1024	sativa	4.4	Uplifted,Happy,Relaxed,Energetic,Creative	Spicy/Herbal,Sage,Woody	1024 is a sativa-dominant hybrid bred in Spain...
3	13-Dawgs	hybrid	4.2	Tingly,Creative,Hungry,Relaxed,Uplifted	Apricot,Citrus,Grapefruit	13 Dawgs is a hybrid of G13 and Chemdawg genet...
4	24K-Gold	hybrid	4.6	Happy,Relaxed,Euphoric,Uplifted,Talkative	Citrus,Earthy,Orange	Also known as Kosher Tangie, 24k Gold is a 60%...

2351 rows x 6 columns

- Strain: object
- Type : object
- Rating: float64
- Effects: object
- Flavor: object
- Description: object

Hypothesis

I want to offer the **perfect range** of cannabis strains in my coffeeshop

1. 6 Stains of each type (sativa, indica & hybrid)
2. high ranking
3. maximum variety in effects and flavors

Bonus:

- Prices
- 1 'Cali' strain per Type



Cleaning (round 1)

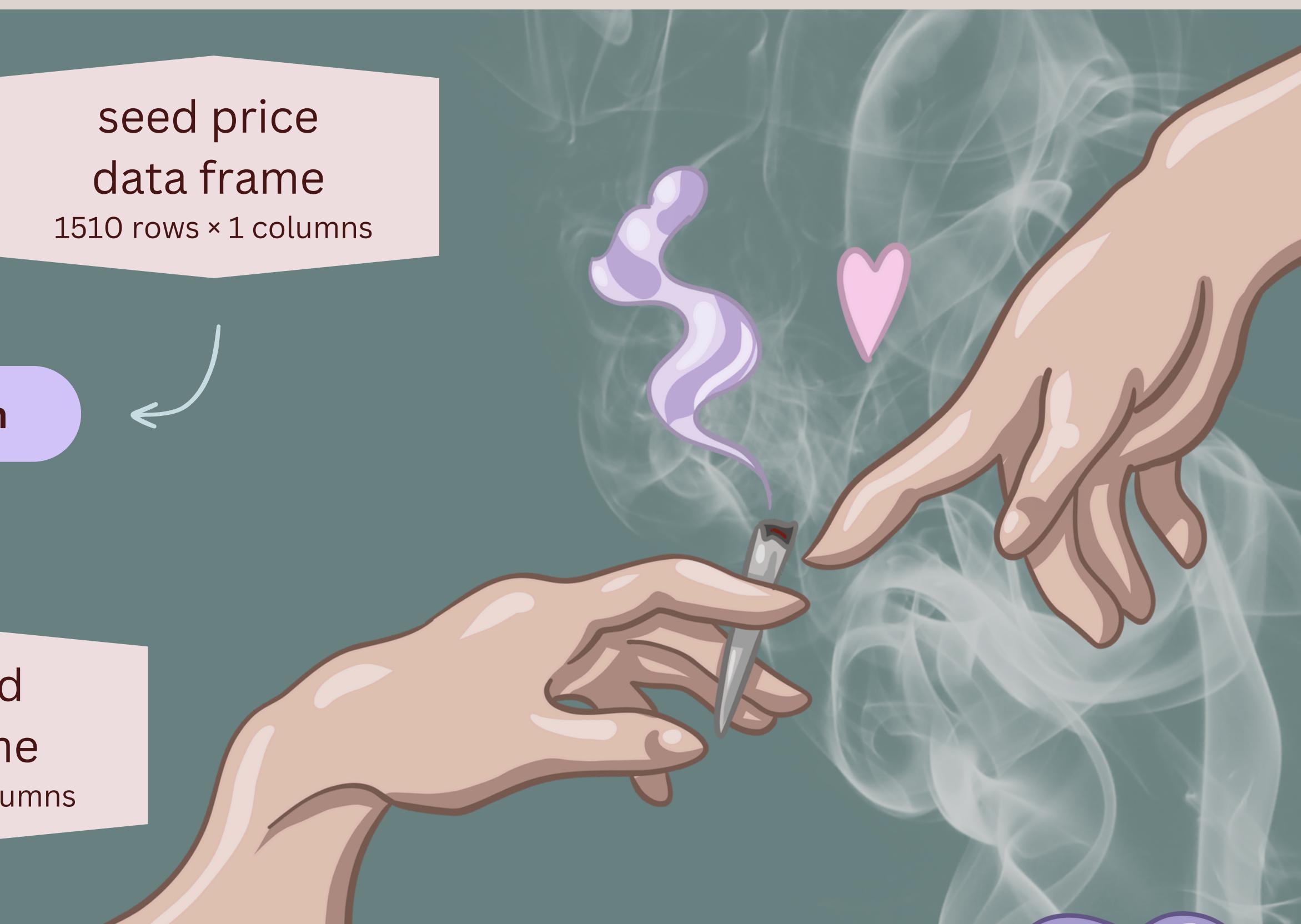
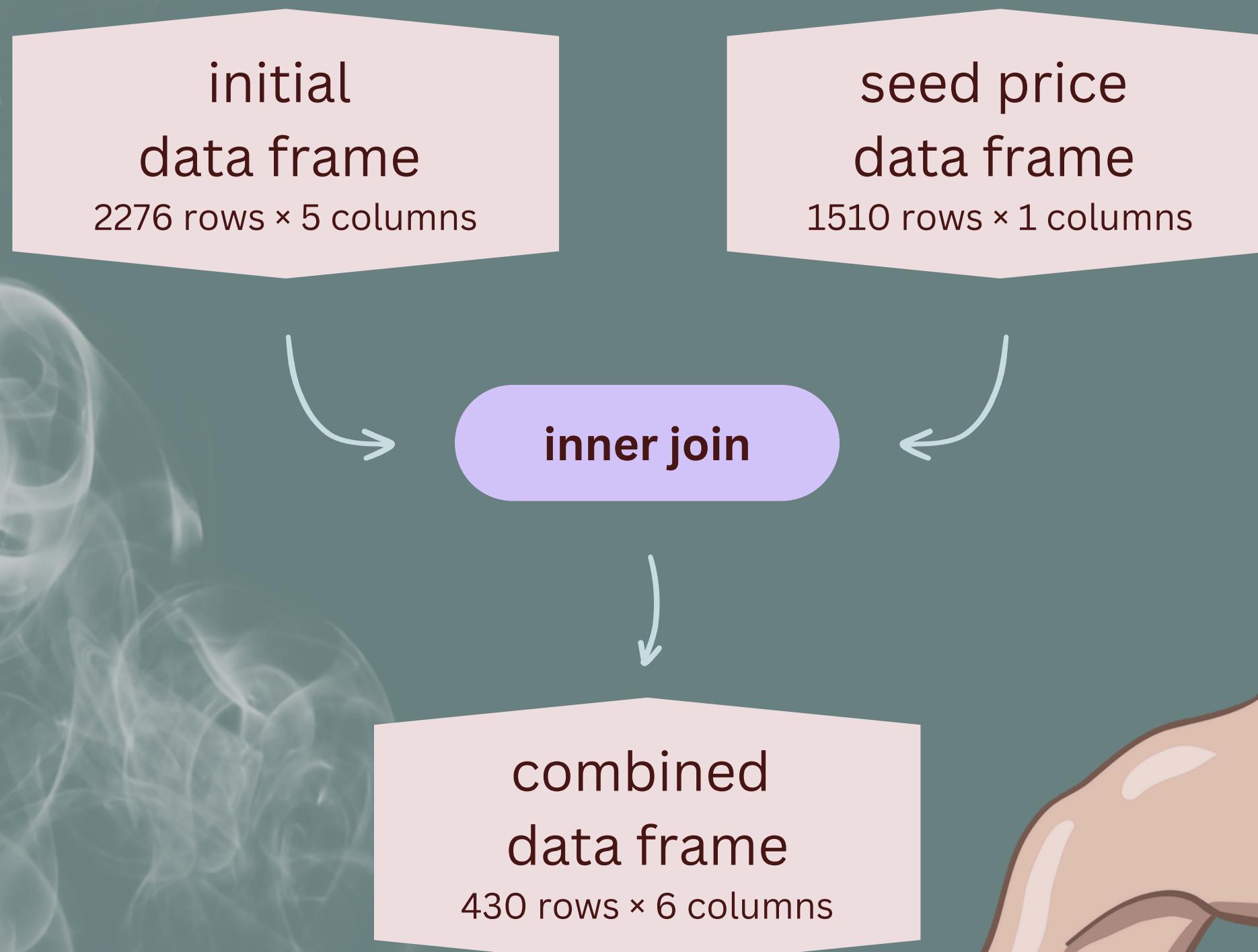
Initial Data:

1. column names & removed duplicates
2. dropped null values (Rating, Flavor, Effect, Description)
3. cleaned “name” column (.lower)
4. set **name as index**

Scraped Data (Seed Retailer Website):

1. column names & removed duplicates
2. cleaned name column
 - Amnesia Haze (Royal Queen Seeds) feminized -> amnesia-haze
3. set **name as index**

Combining the Data Frames



Cleaning (round 2)

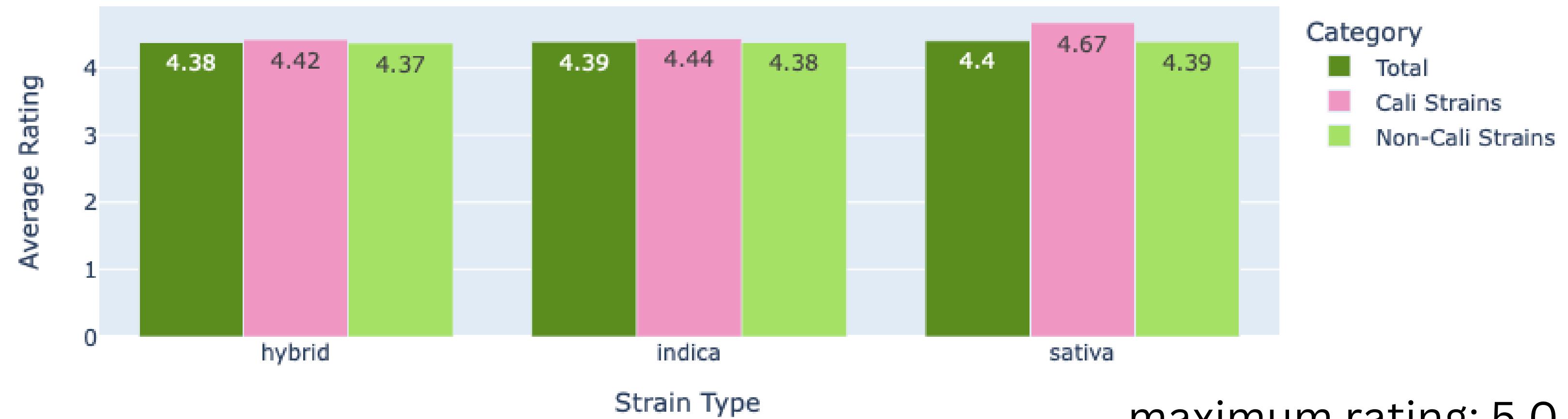
1. Dropped duplicate indices (due to various prices for the same strain)
 - added a step before joining (grouping by name and return median)
2. Encoded flavour
3. Encoded effects
4. Identifying “Cali” strains from description

final
data frame
307 rows × 65 columns

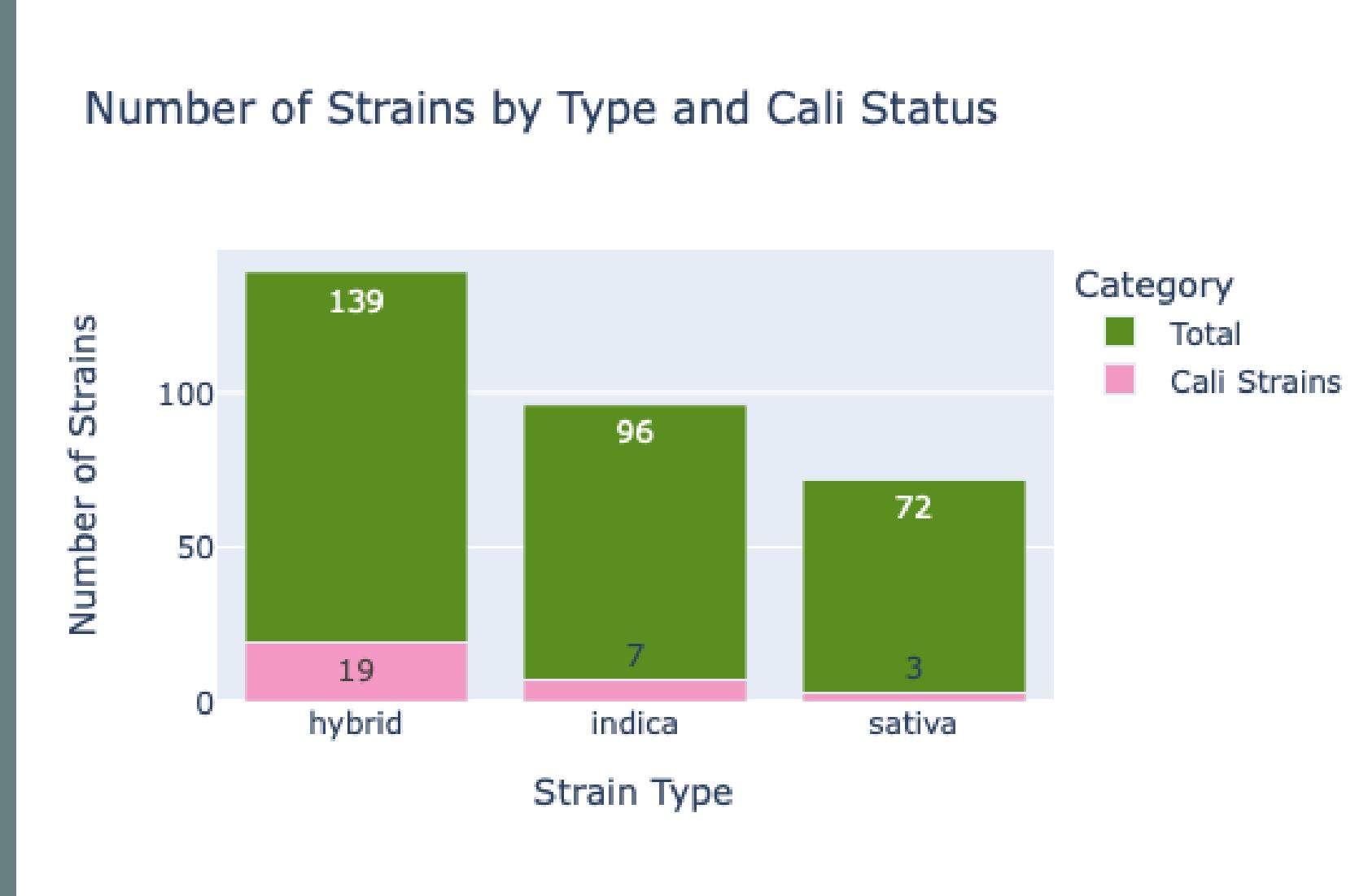
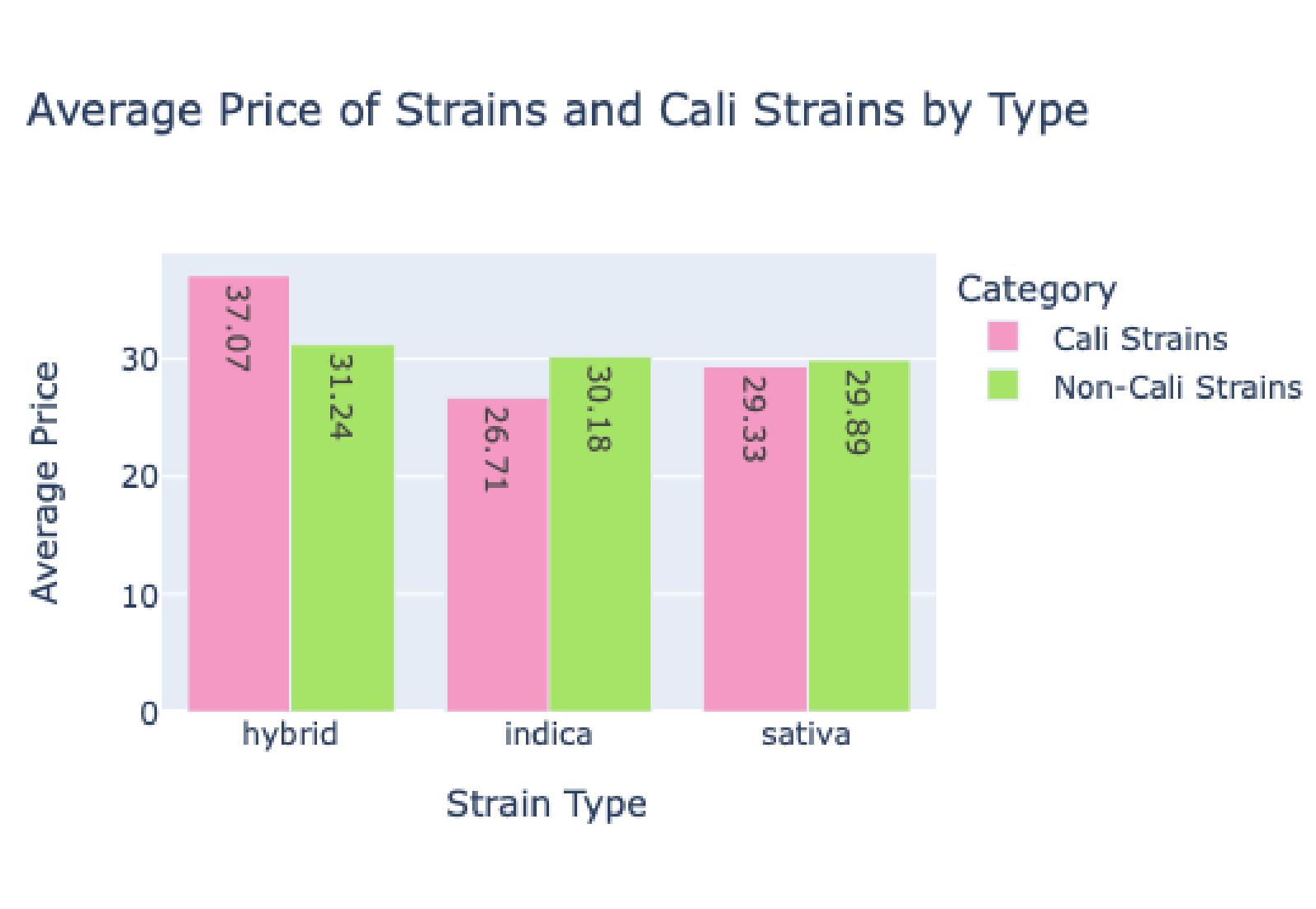


Exploratory Data Analysis

Average Rating of Strains, Cali Strains, and Non-Cali Strains by Type



Exploratory Data Analysis





Chosing Strains

Feature Hierarchy:

1. Rating
2. Effects variaty
3. Flavour variaty

1. Chose the highest ranked Cali Strain for each Type
2. Added **Top 4 ranked** Non-Cali Sativa Strains
3. **5th and 6th place were equally ranked** so I investigated the current distribution of effects and flavors and (since missing effects weren't represented in neither of them I decided to go with the one that added most variaty to flavors)
4. The **6 of the Non-Cali Sativa Indica were rated equally** so I chose first the ones that represented the not yet represented effect 'tingly' and than the 3 that added most variaty to the Flavors
5. Added top 3 **Non-Cali Hybrid Strains** compared 4th - 8th (equally ranked) and chose the ones that added to flavor variaty



Functions

```
def effects_variety(menu):
    # Calculate the sum of effects
    effects_sum = menu.iloc[:, 53:66].sum()

    # Identify missing effects
    missing_effects = effects_sum[effects_sum == 0].index.tolist()

    # Display the total effects and the missing effects
    print("Total effects count for selected strains:")
    print(effects_sum)

    print("\nMissing effects:")
    return missing_effects
```

```
def check_added_effects_variety(df, strains_to_check, effects_to_check):
    # Selecting effects columns for the specified strains
    effects_data = df.loc[strains_to_check, effects_to_check]

    # Display the result
    print(effects_data)
```

```
def flavors_variety(menu):
    # Calculate the sum of flavors for selected strains
    flavors_sum = menu.iloc[:, 7:52].sum()

    # Identify missing effects
    missing_flavors = flavors_sum[flavors_sum == 0].index.tolist()

    # Return the total flavors count and the missing flavors
    return flavors_sum, missing_flavors

def check_added_flavor_variety(strains_to_check, flavors_to_check, df_final):
    # Initialize an empty dictionary to store the results
    missing_flavor_results = {}

    # Iterate over strains to check
    for strain in strains_to_check:
        # Initialize sum of present flavors for the current strain
        sum_present_flavors = 0

        # Check if the strain is present in the index of df_final
        if strain in df_final.index:
            # Iterate over columns in df_final that are also in missing_flavors
            for flavor in flavors_to_check:
                # Check if the flavor column exists in df_final and if its value for the current strain is 1
                if flavor in df_final.columns and df_final.loc[strain, flavor] == 1:
                    # Increment the count of present flavors for the current strain
                    sum_present_flavors += 1

            # Add the sum of present flavors to the dictionary
            missing_flavor_results[strain] = sum_present_flavors

    return missing_flavor_results
```

Menu



Sativa

	rating	effects	flavors
acid-dough	5.0	Talkative, Giggly, Happy, Hungry, Relaxed	Earthy, Woody, Tree Fruit
amnesia-ganja-haze	5.0	Euphoric, Relaxed	Spicy/Herbal, Sweet
congo	5.0	Uplifted, Energetic, Euphoric, Happy, Talkative	Ammonia, Diesel, Sweet
thunder-struck	4.9	Relaxed, Uplifted, Creative, Focused, Talkative	Sweet, Earthy, Tropical
shining-silver-haze	4.8	Talkative, Uplifted, Relaxed, Euphoric, Hungry	Tropical, Mango, Pine

Indica

	rating	effects	flavors
hammerhead	5.0	Sleepy, Aroused, Tingly, Relaxed, Uplifted	Diesel, Berry, Earthy
washing-machine	5.0	Creative, Relaxed, Sleepy, Talkative, Tingly	Earthy, Pungent, Cheese
lucid-bolt	5.0	Giggly, Hungry, Relaxed, Sleepy, Talkative	Blueberry, Berry, Earthy
anubis	5.0	Relaxed, Sleepy, Creative, Focused, Aroused	Grape, Lime, Earthy
fallen-angel	5.0	Hungry, Relaxed, Sleepy, Uplifted, Euphoric	Earthy, Sage, Sweet

Hybrid

	rating	effects	flavors
remo-chemo	5.0	Talkative, Uplifted, Creative, Euphoric, Giggly	Spicy/Herbal, Pepper, Pungent
brooklyn-mango	4.9	Euphoric, Aroused, Uplifted, Happy, Talkative	Mango, Tropical, Citrus
larry-bird-kush	4.9	Relaxed, Euphoric, Energetic, Happy, Uplifted	Sweet, Orange, Earthy
opium	4.8	Happy, Relaxed, Euphoric, Uplifted, Giggly	Earthy, Citrus, Flowery
royal-highness	4.8	Happy, Relaxed, Energetic, Uplifted, Focused	Flowery, Earthy, Cheese

Cali:

	rating	effects
golden-tiger	5.0	Talkative, Uplifted, Energetic, Euphoric, Happy

Cali:

	rating	effects
california-hash-plant	5.0	Uplifted, Creative, Energetic, Giggly, Happy

Cali:

	rating	effects
orange-hill-special	4.9	Happy, Relaxed, Sleepy, Euphoric, Focused

Major Obstacle



Adding Prices to the data frame

- Understanding the Website's Code
- Joining the Data Frames
- Cleaning the names
- Dealing with Strains that had multiple Prices

What would I have done if this was a real business case?

1. Research on customer preference (what's most important to them)
2. Scrape more Retailer Websites
3. For Hybrid strains search for composition in description
4. Negatives effects
5. THC and CBD
6. Competitor Research
7. Price/g (research the cost of growing etc.)





Puff, Puff,
Pass...