

# TRAVEL IN CALIFORNIA

RECOMMENDATION ENGINE



**-TEAM 2 (Weekenders)-**

Elena Harada | Ritu Pardasani | Rucha Kulkarni | Sanjana Thakur



***Got a weekend and  
thinking where to  
go??***



***Weekenders has you  
covered!!!***

# About Weekenders

- A recommendation tool for travel enthusiasts looking to travel within *California*.
- It will take input from users on specific features and recommend activities based on similarity and overall rating.
- Activities will show prices and location.



# Pipeline

1

**Data Sources**  
*Travel Websites*

2

**Data  
Extraction**  
*Beautiful  
soup*

3

**Data  
Cleaning**  
Pandas

4

**Algorithm**  
Collaborative  
Filtering

5

**Visuals**  
Tableau

# Feature Engineering

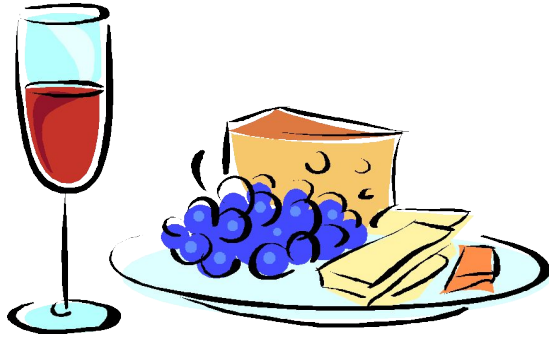
**Activity**



**Region**



# Activity Category



**Food, Wine & Nightlife**



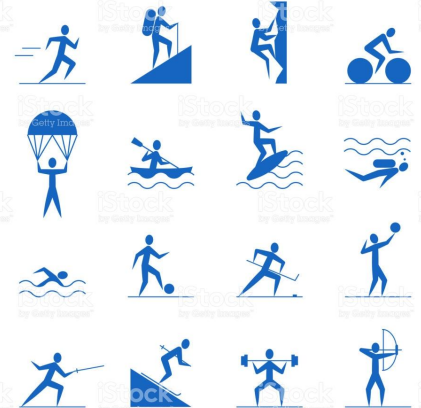
**Water Activities**



**Tours**



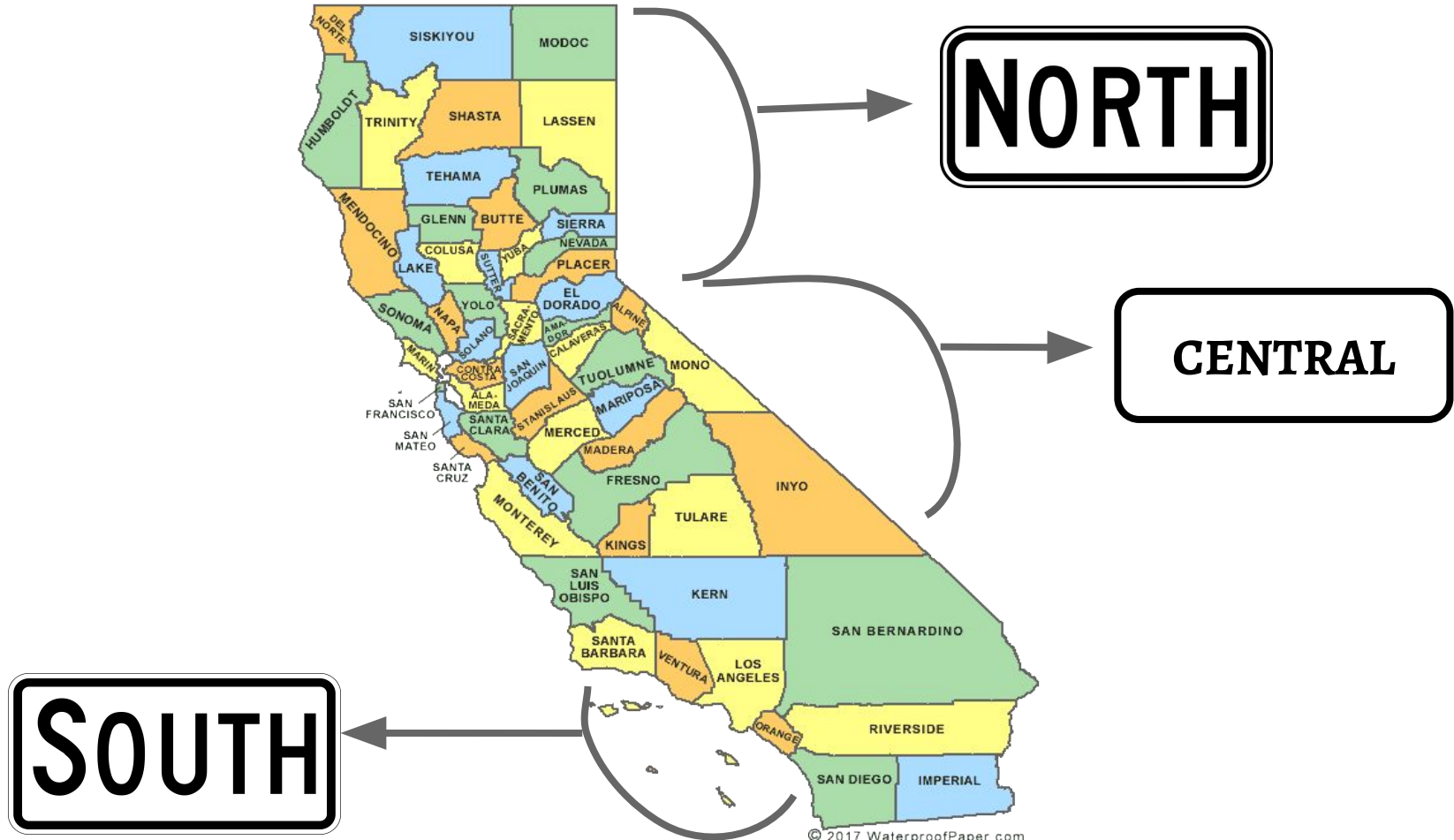
**Walking & Biking**



**Outdoor Activities**



# Region



# Data Source

<https://www.viator.com/California/>

## Categories

- Air, Helicopter & Balloon Tours (59)
- Classes & Workshops (21)
- Cruises, Sailing & Water Tours (142)
- Cultural & Theme Tours (82)
- Day Trips & Excursions (158)
- Food, Wine & Nightlife (328)
- Holiday & Seasonal Tours (15)
- Kid Friendly (73)
- Luxury & Special Occasions (10)
- Multi-day & Extended Tours (36)
- Outdoor Activities (149)
- Private & Custom Tours (113)



Sort by: Most Popular ▼

Page: 1 | 2 | 3 | 4 | 5 | Next →



### Napa and Sonoma Wine Country Tour

★★★★★ 2265 Reviews

San Francisco, California  
8 hours

From USD  
**\$99.00**



### San Francisco Super Saver: Muir Woods & Wine Country w/ optional Gourmet Lunch

★★★★★ 1579 Reviews

San Francisco, California  
9 hours

From USD  
**\$117.00**



### Muir Woods, Giant Redwoods and Sausalito Half-Day Trip

★★★★★ 2479 Reviews

San Francisco, California  
4 hours

From USD  
**\$62.00**



# Data Collection Challenges

- Difficulty scraping data after 2 pages
- Similar list of activities
- Repeated locations or incorrect cities
  - Ex: Santa Ynez may display Solvang or Santa Barbara

# Data Wrangling (Web Scraping)

```
from bs4 import BeautifulSoup
import requests as rq
import pandas as pd
import numpy as np
import re
```

```
url = input("https://www.viator.com/California/d272-ttd")
r = rq.get(url)

data = r.text

soup = BeautifulSoup(data, "html5lib")
```

# Data Wrangling (Tag Region)

```
mapRegion = {  
    'Central': ['Cambria', 'Carmel', 'Los Olivos', 'Monterey', 'Oceano', 'Paso Robles', \  
                'San Luis Obispo', 'Santa Barbara', 'Solvang'],  
    'North': ['Berkeley', 'Cupertino', 'Fish Camp', 'Healdsburg', 'Inverness', 'Jenner', \  
              'Lake Tahoe', 'Livermore', 'Los Gatos', 'Mammoth Lakes', 'Mill Valley', \  
              'Napa', 'Novato', 'Oakhurst', 'Oakland', 'Occidental', 'Petaluma', 'Point Reyes', \  
              'Redding', 'Rohnert Park', 'Sacramento', 'San Francisco', 'San Jose', 'Santa Cruz', 'Santa Rosa', \  
              'Sausalito', 'Sonoma', 'Stockton', 'Tahoe City', 'Truckee'],  
    'South': ['Anaheim', 'Beverly Hills', 'Carlsbad', 'Catalina Island', 'Dana Point', 'Fontana', \  
              'Huntington Beach', 'Joshua Tree', 'La Jolla', 'Laguna Beach', 'Long Beach', 'Los Angeles', \  
              'Malibu', 'Newport Beach', 'Oceanside', 'Palm Desert', 'Palm Springs', 'Pasadena', \  
              'Riverside', 'San Diego', 'Santa Ana', 'Santa Monica', 'Temecula', 'Universal City', \  
              'Venice', 'West Hollywood']  
}
```

```
def tagRegion(x):  
    for k,v in mapRegion.items():  
        if x.strip() in v:  
            return k  
        else:  
            pass
```

# Exported CSV File

Activity	City	Price (USD)	Region	Rating	Water Activities	Outdoor Activities	Walking and Biking	Tours	Food,Wine & Nightlife
10 Hour Santa Cruz Mountains Wine Tasting Tour & Boardwalk Free Time	Santa Cruz	739.99	North	3	0	1	1	1	1
2 Day, 1 Night Kayak Adventure & Camp Out	Stockton	270.17	North	3	1	1	1	1	1
2-Day Alcatraz and Muir Woods With Sonoma Wine and Beer Tour with Aquarium	San Francisco	245	North	4.5	0	0	0	1	1
2-day Alcatraz with Aquarium and Small-Group Muir Woods with Sonoma Wine Tour	San Francisco	252	North	1.5	0	0	0	1	1
2-Day Muir Woods and Wine Country Tour from San Francisco	Mill Valley	344	North	3.5	0	0	0	1	1
2-Day Napa-Sonoma Wine Tour With Alcatraz Night Tour and Aquarium of The Bay	Sonoma	264.5	North	2	0	0	0	1	1

# Region Filtering

```
regionSelection = 'North' #This can be changed based on the user input

def filterRegion(regionSelection):
    if regionSelection == 'North':
        newRegion = raw_data.loc[raw_data['Region'] == 'North']
        return newRegion
    elif regionSelection == 'South':
        newRegion = raw_data.loc[raw_data['Region'] == 'South']
        return newRegion
    elif regionSelection == 'Central':
        newRegion = raw_data.loc[raw_data['Region'] == 'Central']
        return newRegion
    else:
        newRegion = raw_data
        return newRegion
```



# Collaborative Filtering Example - Amazon

## Today's Recommendations For You

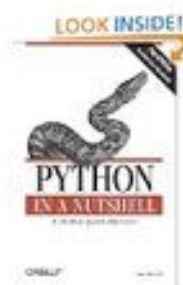
Here's a daily sample of items recommended for you. Click here to [see all recommendations](#).



[Principles of Data Mining \(A...](#) 

by David J....

★★★★☆ (17) \$52.00



[Python in a Nutshell, Secon...](#) 

by Alex Mart...

★★★★☆ (40) \$26.39



[Introductory Statistics wit...](#) 

by Peter Dal...

★★★★☆ (20) \$48.56

# Algorithm - Collaborative Filtering

```
activities_df = raw_data.reset_index(drop=True)
activities_df['Category'] = activities_df.iloc[:, 5:] \
    .apply(lambda x: ','.join(x.dropna().astype(int).astype(str)), axis=1)
activities_df['Category'] = activities_df['Category'] \
    .apply(lambda x: list(map(int, x.split(','))))
activities_tuple = [tuple(x) for x in activities_df.iloc[:, 0:6].values]
activities_dict = {i:j for i,j in enumerate(activities_tuple)}
```

```
activities_dict[900]
```

```
('Sonoma Valley Wine Tour from San Francisco',
 'Sonoma',
 '139',
 'North',
 4.0,
 [0, 0, 0, 1, 1])
```

# Algorithm - Collaborative Filtering

```
from scipy import spatial

def ComputeDistance(a, b):
    # a[5] = category
    catA = a[5]
    catB = b[5]
    catDistance = spatial.distance.cosine(catA, catB)
    # a[4] = rating
    ratingA = a[4]
    ratingB = b[4]
    ratingDistance = abs(ratingA - ratingB)
    return (catDistance + ratingDistance)
```

# Algorithm - Collaborative Filtering

```
import operator

def getNeighbors(activityName, K):
    distances = []
    for activity in activities_dict:
        if (activity != activityName):
            dist = ComputeDistance(activities_dict[activityName], activities_dict[activity])
            distances.append((activity, dist))
    distances.sort(key=operator.itemgetter(1))

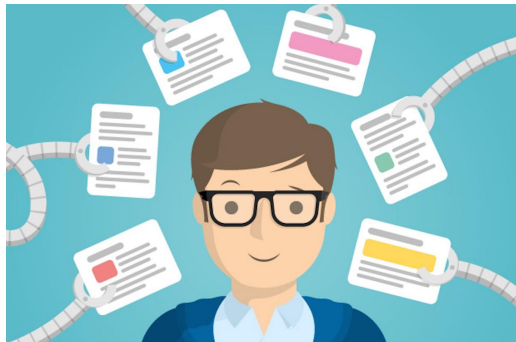
    neighbors = []
    for x in range(K):
        neighbors.append(distances[x][0])
    return neighbors
```

# Algorithm - Collaborative Filtering

```
K = 5

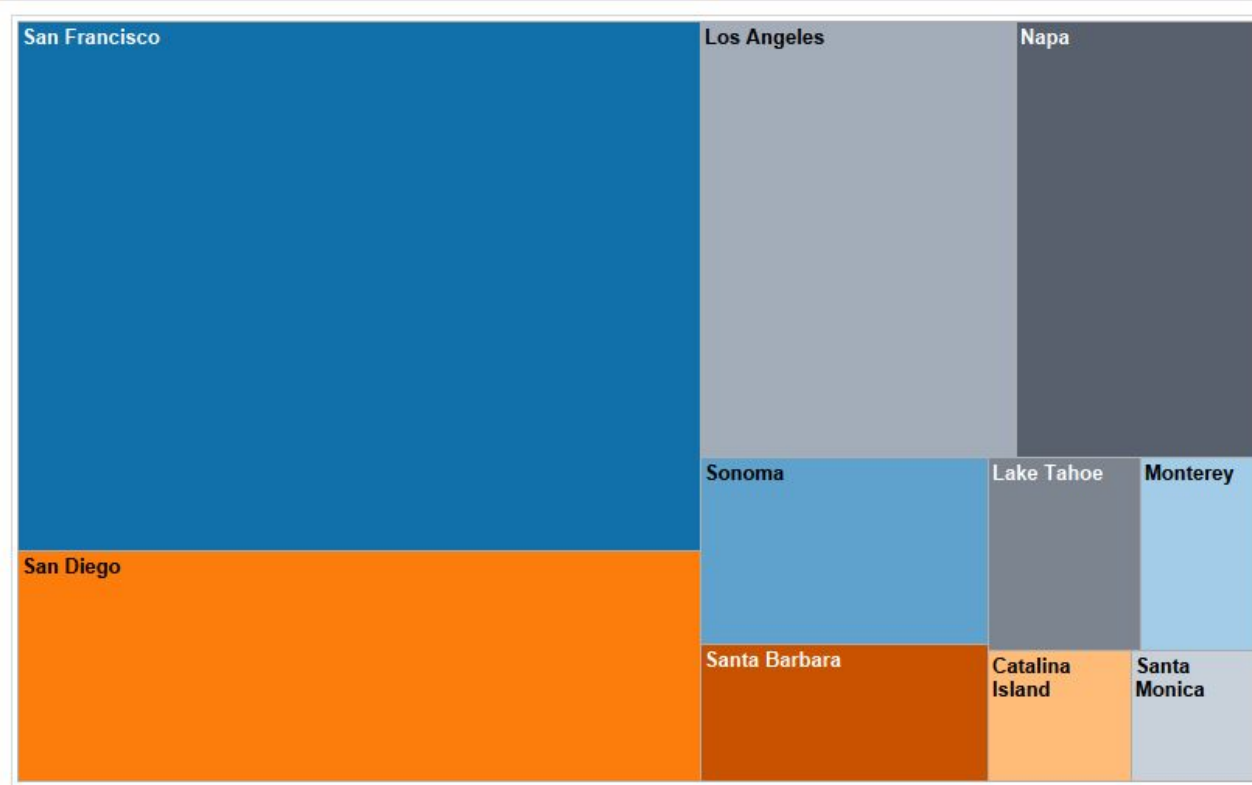
neighbors = getNeighbors(900, K)
for y in neighbors:
    print(activities_dict[y][0] + ", $" + str(activities_dict[y][2]) \
          + ", " + str(activities_dict[y][1]))
```

6 Hour Dessert Wine Tour - Napa Valley, \$480, San Francisco  
6 Hour Private Wine Tasting Tour with Pre-set Stops, \$439.99, Napa  
8 Hour Dry Creek Valley Wine Tasting Tour, \$549.95, Sonoma  
8 Hour Santa Cruz Mountains Wine Tasting Tour from San Francisco, \$599.99, Santa Cruz  
8 Hour South Bay Wine Tasting Tour from San Francisco, \$539.88, San Jose



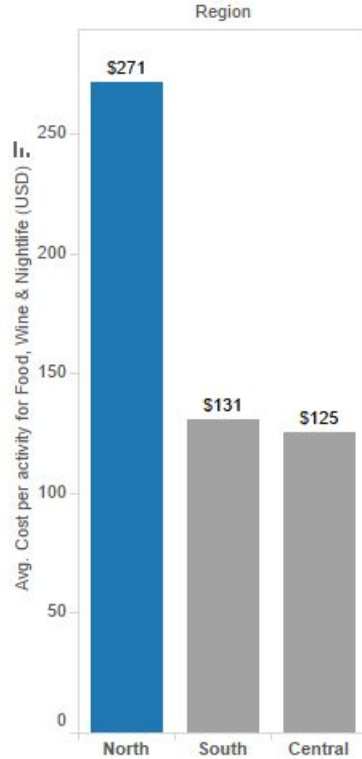


# Tableau Visualizations

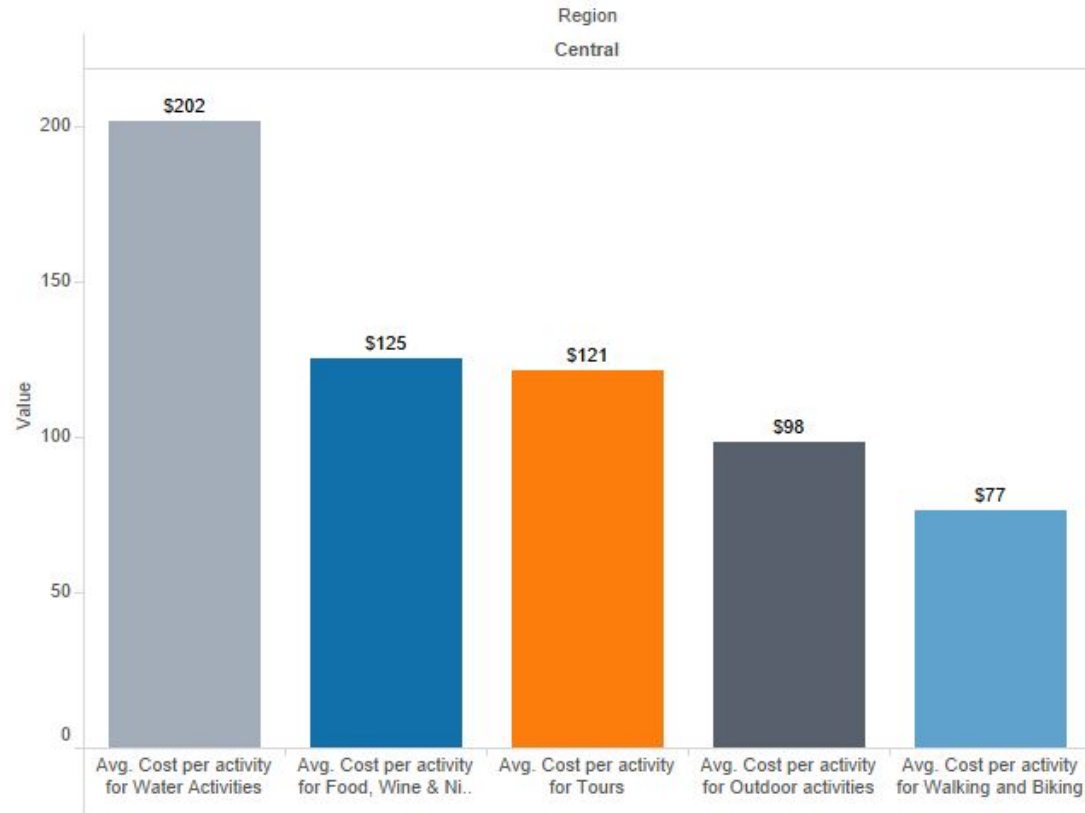


# Tableau Visualizations

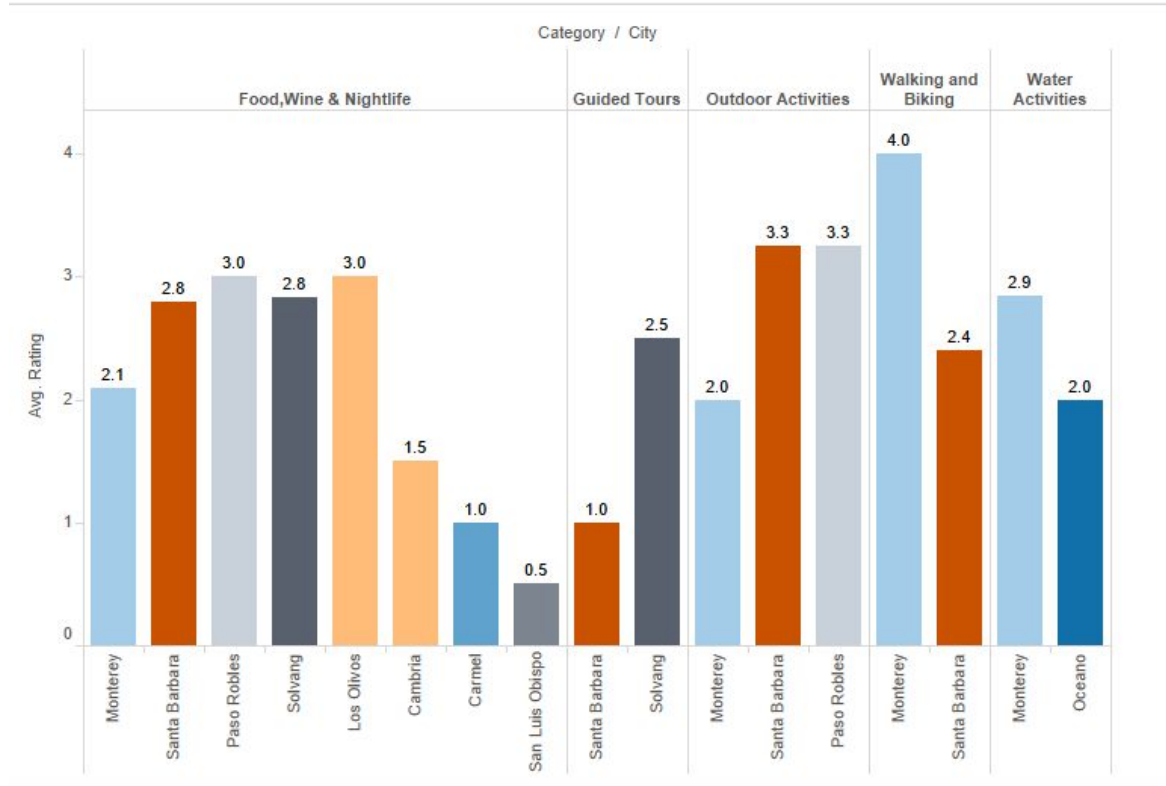
Most expensive region for average cost for the Food, Wine and Nightlife category



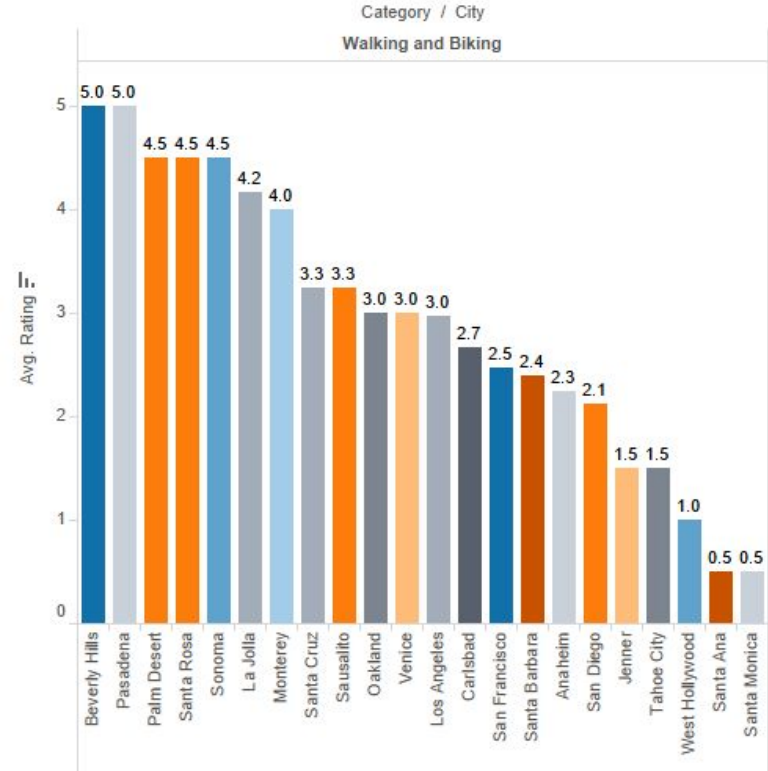
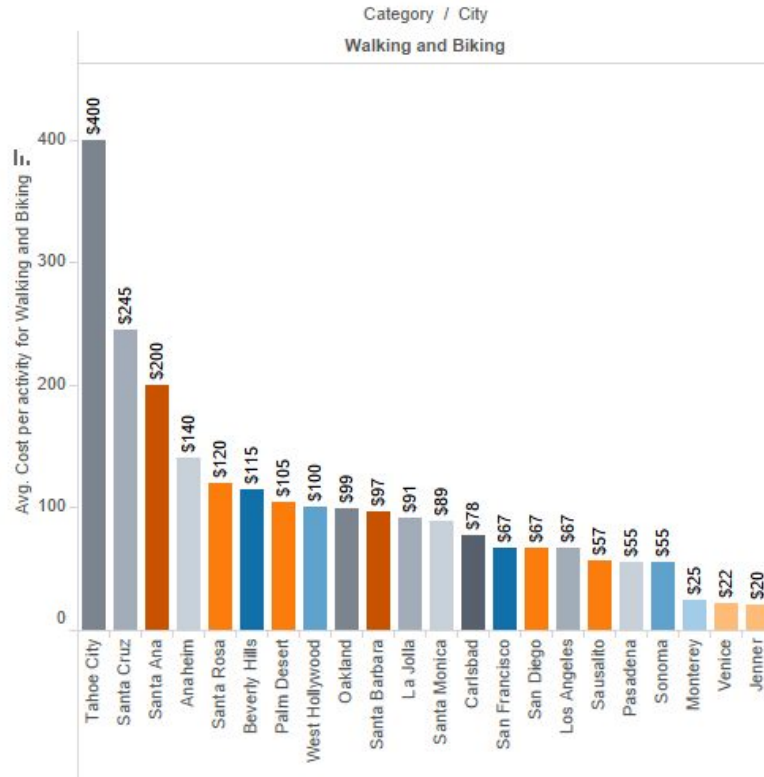
# Tableau Visualizations



# Tableau Visualizations



# Tableau Visualizations





# Recommendation Website

[Home](#)[Explore by category of activities](#)[Explore by region of activities](#)[More](#)

## *The Weekenders*

*Your best guide to make the most of  
California!*

*At The Weekenders, we don't provide the typical search engine results you get. We work relentlessly to get you the ratings of the activities you can plan, the cost that you can expect and the best you can plan for your upcoming weekend getaway!*

[Explore by category of activity](#)[Explore by region of activity](#)

# Recommendation Website (LINKS)

Explore by category of activities

Walking and Biking

Water Activities

Food, Wine and Nightlife

Tours

Outdoor activities

Explore by region of activities

Central California

Northern California

Southern California

More

Fun Facts

Contact

**NEW TO CALIFORNIA? HERE ARE THE MOST HAPPENING CITIES!**

**MOST EXPENSIVE REGION FOR AVERAGE COST FOR THE FOOD, WINE AND NIGHTLIFE**

# Recommendation Website (LINKS)

Wine Tasting ▼

North California ▼

Submit

Activity	City	Price (USD)	Region
6 Hour Dessert Wine Tour - Napa Valley	San Francisco	480	North
6 Hour Private Wine Tasting Tour with Pre-set Stops	Napa	439.99	North
8 Hour Dry Creek Valley Wine Tasting Tour	Sonoma	549.95	North
8 Hour Santa Cruz Mountains Wine Tasting Tour from San Francisco	Santa Cruz	599.99	North
8 Hour South Bay Wine Tasting Tour from San Francisco	San Jose	539.88	North

Back to search

# Future Versions

- **Include:**
  - Price Range
  - Trip Duration
  - Number of Travelers
- **Expand to all states in USA**

THANK YOU