

# Coffee Shops Data Analysis

**Presented by Coffee Beans**

**Name**

**ID**

**Bhone Myint San**

**6520225**

**Myo Ko Zin**

**6520227**

**Noel Paing Oak Soe**

**6530183**

**Chalida Hiranmateekul**

**6520180**



# Domain and Question

**Domain: Coffee shops in Bangkok**

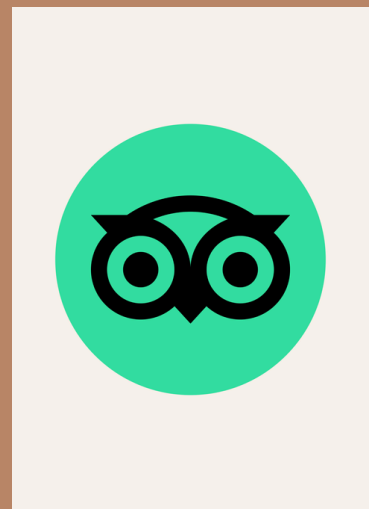
**Questions:**

- 1. Which coffee shops have the highest average ratings?**
- 2. Are there any trends in reviews based on user location?**
- 3. Which coffee shops receive the most reviews?**
- 4. What is the correlation between the length of the review text and the rating given?**



# Scraping Data

Data was scraped from  
TripAdvisor.com using Apify



## Attributes:

1. CoffeeShop,
2. Username,
3. Title, Rating,
4. PublishedDate,
5. ReviewText,
6. UserFrom,
7. Location.



# Data Preview



CoffeeShop	Username	Title	Rating	publishedDate	ReviewText	UserFrom	Location
Size S Coffee & Bakery	michael t	Great coffee on Rama 4 area	5	2024-06-04	Great coffee in Rama 4 area. It's a small shop...	Chicago, Illinois	5/25 Soi Ngam Du Phli
Size S Coffee & Bakery	Stef B	My favorite coffee place in Bangkok	5	2023-10-22	My favorite coffee is Bangkok. I like to order...	Berlin, Germany	5/25 Soi Ngam Du Phli
Size S Coffee & Bakery	MollyMonchaya	Size S portion, Size L for price	3	2020-02-26	It is a nice and cosy place,,just 100 m from m...	NaN	5/25 Soi Ngam Du Phli
Size S Coffee & Bakery	Travellingflamingo	Unexpected catch!	5	2019-06-09	Passed by this unassuming little place going t...	Panglao Island, Philippines	5/25 Soi Ngam Du Phli
Size S Coffee & Bakery	Tyke T	Excellent	5	2019-02-01	Very nice - but small - cafe, a bit hidden.\nG...	Malmö, Sweden	5/25 Soi Ngam Du Phli

Shape of the dataframe: (7057, 8)



# Data Cleaning Methods Used

- **Dropping Unnecessary Columns**
- **Renaming and Reordering the Columns**
- **Dropping Null Values**



#	Column	Non-Null	Count	Dtype
0	CoffeeShop	7057	non-null	object
1	Username	7026	non-null	object
2	Title	7053	non-null	object
3	Rating	7057	non-null	int64
4	publishedDate	7057	non-null	object
5	ReviewText	7057	non-null	object
6	UserFrom	5146	non-null	object
7	Location	7057	non-null	object

Missing values:

CoffeeShop	0
Username	31
Title	4
Rating	0
publishedDate	0
ReviewText	0
UserFrom	1911
Location	0





# Descriptive Analysis

Descriptive statistics:

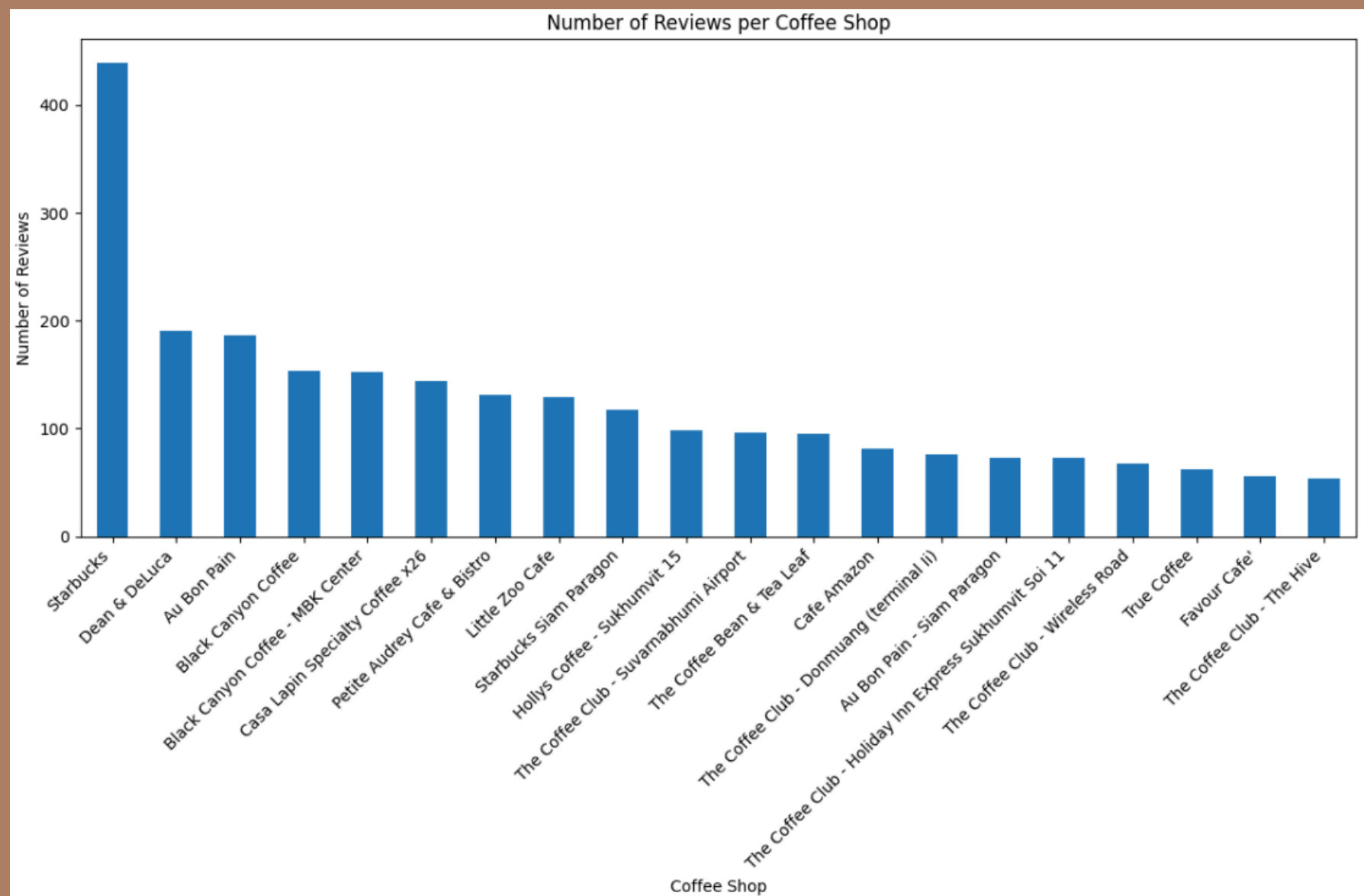
	Rating
count	7057.000000
mean	4.088565
std	1.045835
min	1.000000
25%	4.000000
50%	4.000000
75%	5.000000
max	5.000000



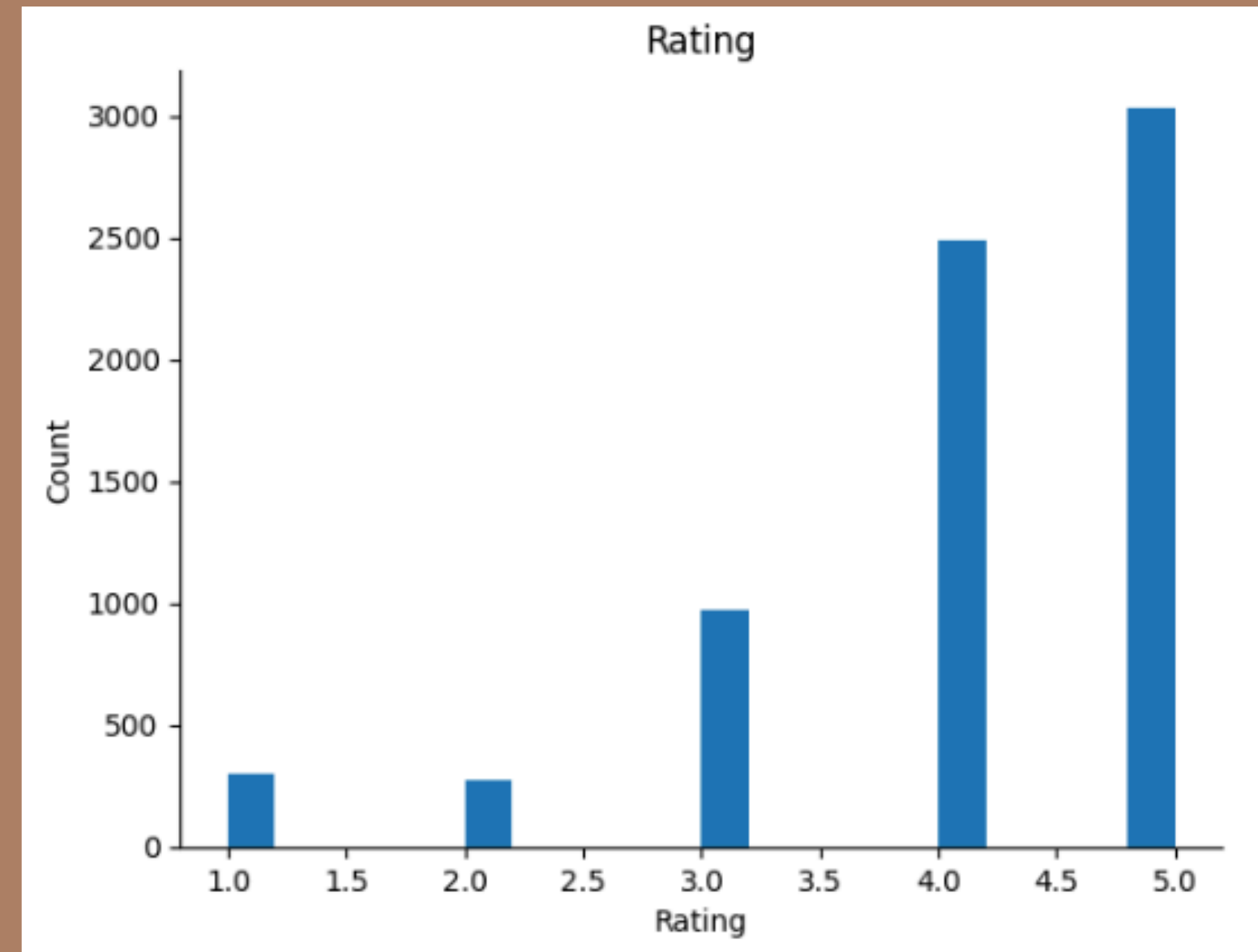
[https://colab.research.google.com/drive/1g0zmiZtvcrcxrA3fuSXUf3IwjWCAvApI#scrollTo=MR\\_eIBufkSu5](https://colab.research.google.com/drive/1g0zmiZtvcrcxrA3fuSXUf3IwjWCAvApI#scrollTo=MR_eIBufkSu5)



# Data Exploration



Coffee shops and numbers of their reviews



Number of rating (1 – 5)



# Answering the Questions

**Which coffee shops have the highest average ratings?**

```
☞ Coffee shops with the highest average ratings:  
CoffeeShop  
naam hostel n cafe      5.0  
Budhtri Coffee          5.0  
Embassy Coffee          5.0  
Foxhole BKK             5.0  
Kaarom Specialty Coffee 5.0  
Name: Rating, dtype: float64
```

**Are there any trends in reviews based on user location?**

```
UserFrom  
Gerik, Malaysia      5.0  
Royan, France        5.0  
Heerlen, The Netherlands 5.0  
Heidelberg, Germany  5.0  
Heimiswil, Switzerland 5.0  
Helsingoer Municipality, Denmark 5.0  
Hendaye, France      5.0  
Henderson, Nevada    5.0  
Hengelo, The Netherlands 5.0  
Hercules, California  5.0
```



# Answering the Questions

**Which coffee shops receive the most reviews?**

```
Coffee shops receiving the most reviews:  
CoffeeShop  
Starbucks          439  
Dean & DeLuca     191  
Au Bon Pain       187  
Black Canyon Coffee 154  
Black Canyon Coffee - MBK Center 153  
Casa Lapin Specialty Coffee x26 144  
Petite Audrey Cafe & Bistro 131  
Little Zoo Cafe    129  
Starbucks Siam Paragon 117  
Hollys Coffee - Sukhumvit 15 98  
Name: count, dtype: int64
```

**What is the correlation between the length of the review text and the rating given?**

```
Correlation between review length and rating: -0.14996712778148633
```



# **What's for future?**

- **Sentiment Analysis**
- **Train a model**
- **Predict User's rating**

**Presented by Coffee Beans**

**Thank you  
very much!**