

# Analysis Data File:

## Combined Rating and Sentiment Analysis Dataset

### Unit of Observation

Each row in this dataset represents a unique user's review, meaning that each row corresponds to a specific review of a movie theater chain.

---

## Variable Definitions and Summaries

### 1. Theater Name

- **Definition:** The name of the movie theater brand.
- **Source Data:** Google Reviews
- **Processing Steps:** Taken directly from Google Reviews using Phantom Web Scraper.
- **Missing Observations:** N/A
- **Summary Statistics:**
  - Alamo Drafthouse: 1000
  - Violet Crown: 941
  - Regal: 1000

### 2. Review Date

- **Definition:** The date that a user left a review for a particular movie theater.
- **Source Data:** Google Reviews
- **Processing Steps:** Taken directly from Google Reviews using Phantom Web Scraper.
- **Missing Observations:** The phantom web scraper has a limit of 1000 reviews per business. While the Violet Crown only had 942, Regal had 1614, and Alamo had 2346. This means a total of 1960 reviews were unable to be counted.
- **Summary Statistics:**

Year	Month	Total Reviews
2015	Jul	2
	Aug	2
	Sep	2

	Oct	2
	Nov	2
	Dec	2
<b>2015 Total</b>		<b>12</b>
<b>2016</b>	Jan	4
	Feb	10
	Mar	10
	Apr	10
	May	10
	Jun	10
	Jul	10
	Aug	10
	Sep	10
	Oct	10
	Nov	10
	Dec	10
<b>2016 Total</b>		<b>114</b>
<b>2017</b>	Jan	13
	Feb	18
	Mar	18
	Apr	18
	May	18
	Jun	18
	Jul	18
	Aug	18
	Sep	18
	Oct	18

	Nov	18
	Dec	18
<b>2017 Total</b>		<b>211</b>
<b>2018</b>	Jan	9
	Feb	26
	Mar	26
	Apr	26
	May	26
	Jun	26
	Jul	26
	Aug	26
	Sep	26
	Oct	26
	Nov	26
	Dec	26
<b>2018 Total</b>		<b>295</b>
<b>2019</b>	Jan	44
	Feb	53
	Mar	53
	Apr	53
	May	56
	Jun	56
	Jul	56
	Aug	56
	Sep	56
	Oct	56
	Nov	56

	Dec	56
<b>2019 Total</b>		<b>651</b>
<b>2020</b>	Jan	4
	Feb	14
	Mar	19
	Apr	19
	May	19
	Jun	19
	Jul	19
	Aug	19
	Sep	19
	Oct	19
	Nov	19
	Dec	19
<b>2020 Total</b>		<b>208</b>
<b>2021</b>	Jan	9
	Feb	24
	Mar	26
	Apr	28
	May	28
	Jun	28
	Jul	28
	Aug	28
	Sep	28
	Oct	28
	Nov	28
	Dec	28

<b>2021 Total</b>		<b>311</b>
<b>2022</b>	Jan	21
	Feb	32
	Mar	37
	Apr	37
	May	37
	Jun	37
	Jul	37
	Aug	37
	Sep	37
	Oct	37
	Nov	37
	Dec	37
<b>2022 Total</b>		<b>423</b>
<b>2023</b>	Jan	23
	Feb	32
	Mar	35
	Apr	35
	May	35
	Jun	35
	Jul	35
	Aug	35
	Sep	35
	Oct	35
	Nov	35
	Dec	35
<b>2023 Total</b>		<b>405</b>

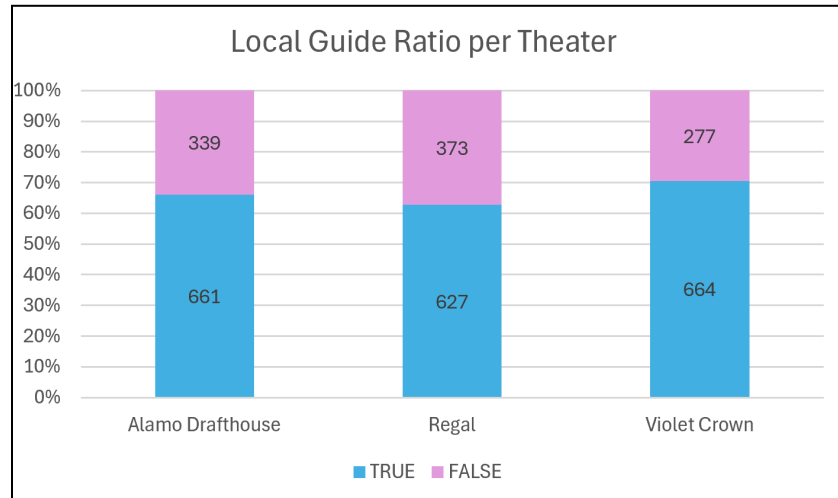
<b>2024</b>	Jan	3
	Feb	22
	Mar	27
	Apr	22
	May	24
	Jun	21
	Jul	39
	Aug	36
	Sep	27
	Oct	18
	Nov	20
	Dec	29
<b>2024 Total</b>		<b>288</b>
<b>2025</b>	Jan	23
<b>2025 Total</b>		<b>23</b>

### 3. Local Guide

- **Definition:** A binary data column that checks whether or not the user who left a review was a local guide. A local guide is someone who contributes to many reviews for local businesses.
- **Source Data:** Google Reviews
- **Processing Steps:** Taken directly from Google Reviews using Phantom Web Scraper.
- **Missing Observations:** A total of 1960 reviews were unable to be counted.
- **Summary Statistics:**

Theater	Local Guides	Non-Local Guides
Violet Crown	664	277
Alamo Drafthouse	661	339
Regal Cinema	627	373

- **Visualization:**

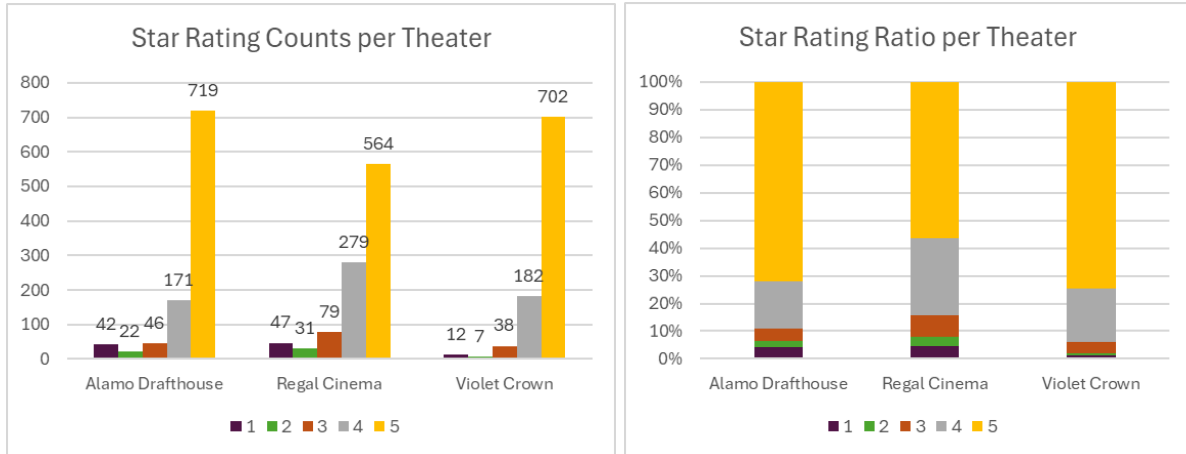


#### 4. Star Rating

- **Definition:** A scale of how much a user liked their experience at the movie theater, with 1 being the least positive experience, and 5 being the highest positive experience.
- **Source Data:** Google Reviews
- **Processing Steps:** Taken directly from Google Reviews using Phantom Web Scraper.
- **Missing Observations:** A total of 1960 reviews were unable to be counted.
- **Frequency Table:**

Theater	1-star	2-star	3-star	4-star	5-star
Alamo Drafthouse	42	22	46	171	719
Regal Cinema	47	31	79	279	564
Violet Crown	12	7	38	182	702

- **Visualization:**



## 5. Review Content

- **Definition:** The written review of a user left for the movie theater. This is all text.
- **Source Data:** Google Reviews
- **Processing Steps:** Taken directly from Google Reviews using Phantom Web Scraper.
- **Missing Observations:** A total of 1960 reviews were unable to be counted.
- **Visualization:** N/A

## 5. Sentiment Polarity

- **Definition:** The Sentiment Polarity refers to the score of how positive or negative a text review was. A score closer to 1 is considered positive, while a score closer to -1 is considered negative. Any score close to 0 is considered neutral. [-1, 1]
- **Source Data:** Python Sentiment Analysis of Scraped Reviews
- **Processing Steps:** These are taken from the sentiment analysis done on our Jupyter Notebook file using TextBlob.
- **Missing Observations:** A total of 1960 reviews were unable to be counted, and some reviews had a rating, but no review content, leading to no sentiment score given. A total of 1350 reviews had a star rating, but no review content.
- **Summarized Data:**

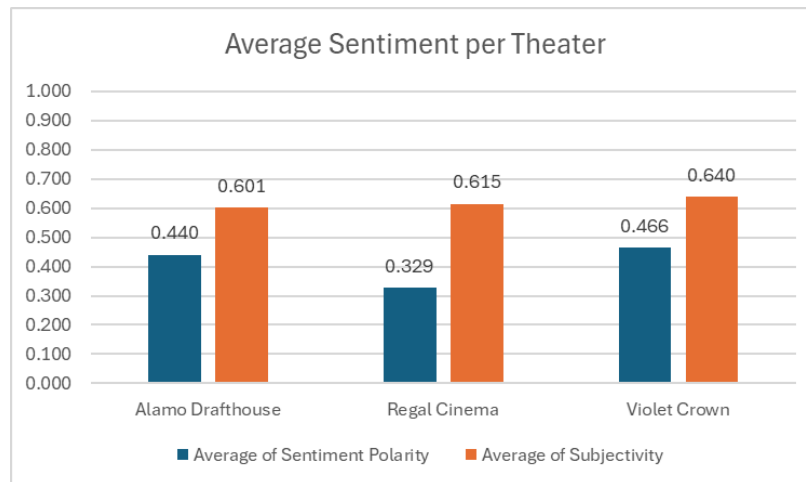
Theater Name	Average of Sentiment Polarity
Alamo Drafthouse	0.440
Regal Cinema	0.329



**Violet Crown**

**0.466**

- **Visualization:** Column Scatterplot of Sentiment Polarity [Blue Bars]



## 5. Subjectivity

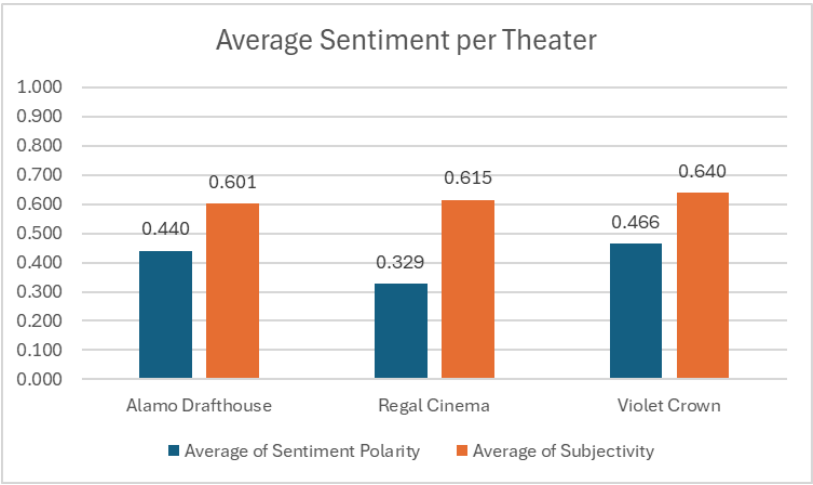
- **Definition:** Subjectivity refers to the score of how opinionated a text review was. A score closer to 1 is considered highly subjective (i.e. more opinionated), while a score closer to 0 is considered less subjective (i.e. more factual). [0,1]
- **Source Data:** Python Sentiment Analysis of Scraped Reviews
- **Processing Steps:** These are taken from the sentiment analysis done on our Jupyter Notebook file using TextBlob.
- **Missing Observations:** A total of 1960 reviews were unable to be counted. No sentiment score was given for 1350 reviews.
- **Summarized Data:**

Theater Name	Average of Subjectivity
Alamo Drafthouse	0.601
Regal Cinema	0.615

Violet Crown

0.640

- **Visualization:** Column Scatterplot of Subjectivity [Orange Bars]



---