

## AGENDA



Business Use Case



Data Preparation



Database Design



Reporting and Visualization

## Executive Summary

### DATA DRIVEN STRATEGY

Multi-dimensional insights for business owners to grow business

#### YELP DATA

Most widely used restaurant review platform



YELP REVIEWS



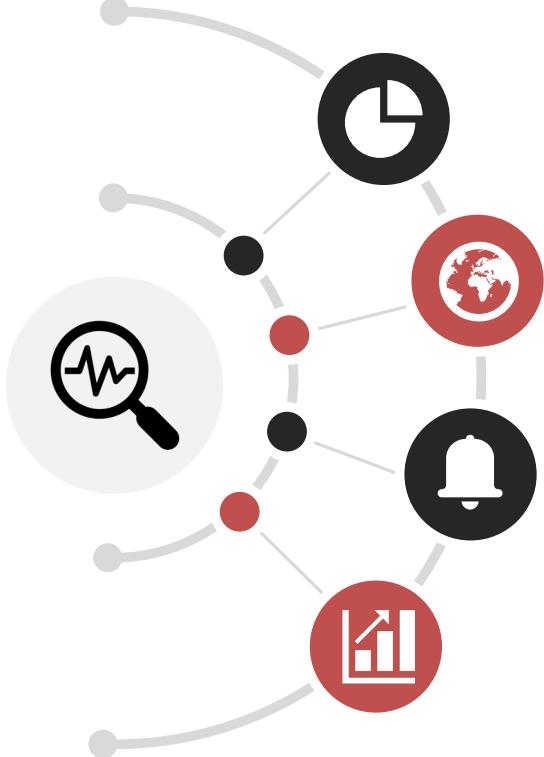
GROWTH



SOCIAL MEDIA

#### TWITTER DATA

Provide additional dimension for consumer insights



### **YELP DATA MANAGEMENT**

Generate descriptive analysis about yelp restaurant data

### **SOCIAL MEDIA & DEMOGRAPHIC DATA**

Create multidimensional yelp database and cross reference with secondary data source (i.e. twitter and IRS)

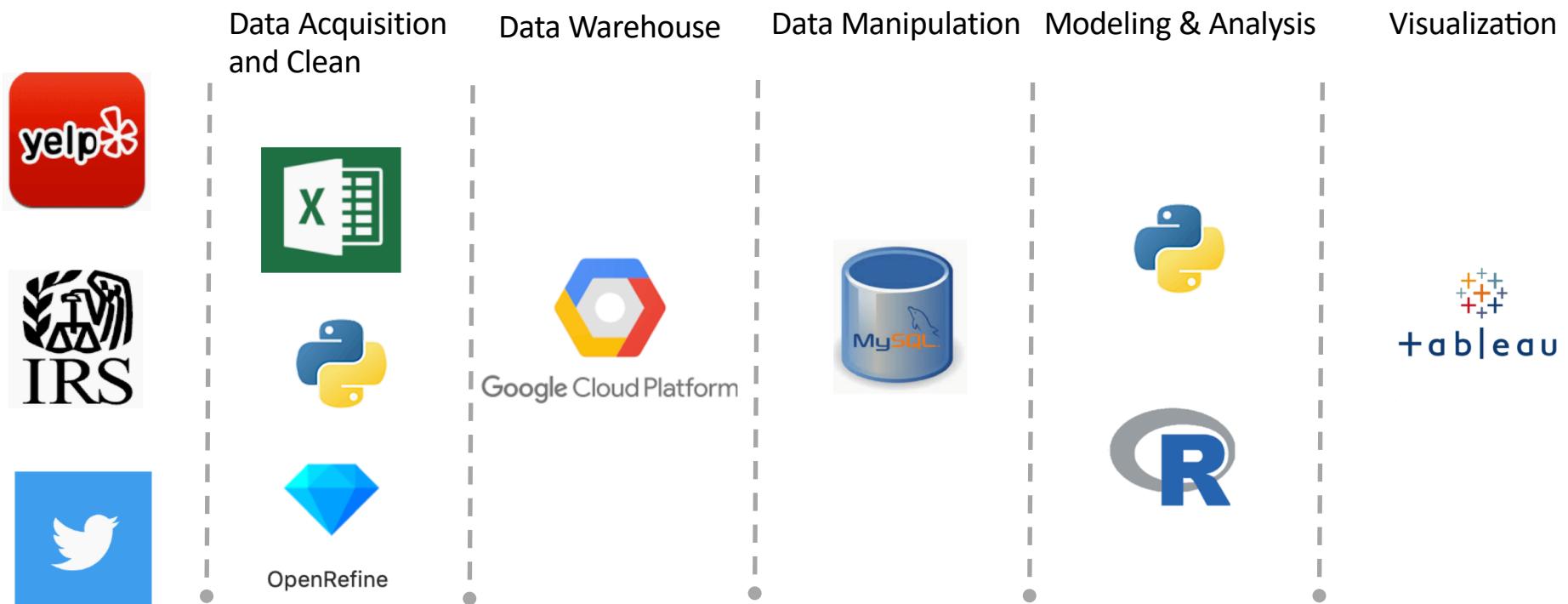
### **USER-FRIENDLY DASHBOARDING**

Provide user-friendly dashboard for data tracking and visualization

### **MKT & PR STRATEGY**

Generate data-driven insights for marketing and PR strategies

## Data Preparation



## ETL-Process



Extraction	<u><a href="#">Yelp Open Dataset</a></u> - 580k reviews - 32k business	Tweepy Python Package - 80 restaurants accounts - 16K tweets	<u><a href="#">IRS household income dataset</a></u>
Transformation	- Only included those open restaurants and related reviews - Excluded unrelated attributes, such as 'HairSpecializesIn'	- Transform and read JSON file into one flat file to future storage in SQL - Recode the tweets_id to easy reference	- Excluded unnecessary columns - Transform Geolocation information to match other 2 data source
Loading	Load the cleaned datasets in google cloud storage		

## ETL issues faced and how we addressed

### Issues

- Data volume and loading
- Multiple platforms transition

### Solutions

- Load in separate batches
- Index creation to facilitate query efficiency
- GCP cloud base

## Database Design

OLAP system for analytics, BI and decision supporting purpose



### Conceptual Modeling



### Logical Modeling



### Physical Modeling

- Entity Definition:
  - Yelp Merchants
  - Yelp Consumers
  - Yelp Reviews
  - Yelp Checkins
  - Twitter Users
  - Tweets
  - States
  - etc.
- Entity Relationship & Cardinality
  - Merchants v. Consumers: N:M
  - Reviews v. Merchants: N:1
  - Reviews v. Consumers: N:1
  - Merchants v. Twitter Users: 1:1
  - Twitter Users v. Tweets: 1:N
  - Twitter Users v. Users: 1:N
  - etc.

## Database Design

OLAP system for analytics, BI and decision supporting purpose



### Conceptual Modeling



### Logical Modeling



### Physical Modeling

- Table Definition:
  - Yelp Merchants (identified by Business\_id)
  - Yelp Consumers (identified by User\_id)
  - Yelp Reviews (identified by Review\_id)
  - Twitter Users (identified by Account)
  - Tweets (identified by Tweet\_id)
  - ...
- Restriction & Foreign Keys
- OLAP: Normalization & Non-normalization

## Database Design

OLAP system for analytics, BI and decision supporting purpose



### Conceptual Modeling



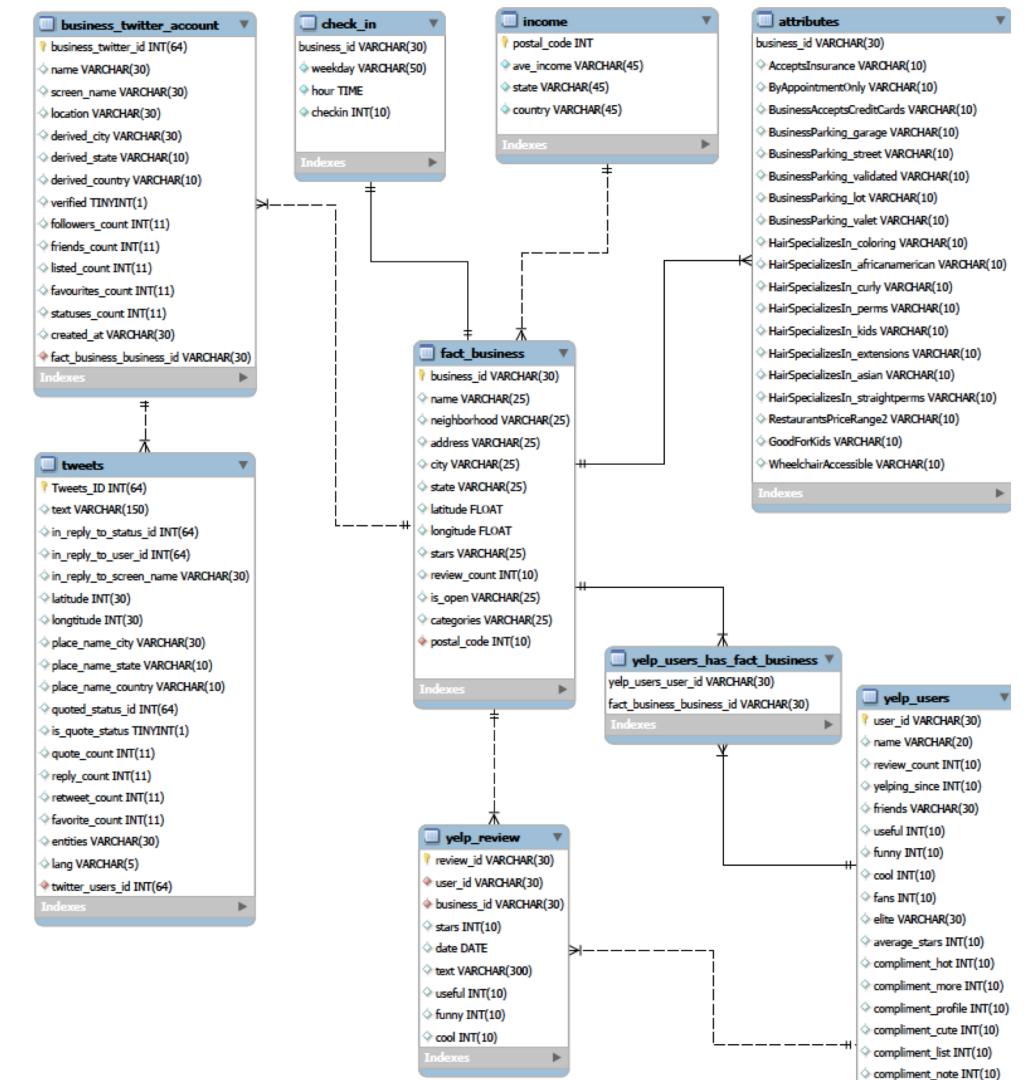
### Logical Modeling



### Physical Modeling

- Table definition:
  - Yelp Merchants (identified by Business\_id)
  - Yelp Consumers (identified by User\_id)
  - Yelp Reviews (identified by Review\_id)
  - Twitter Users (identified by Account)
  - Tweets (identified by Tweet\_id)
  - ...
- Restriction & Foreign Keys
- OLAP: Normalization & Non-normalization
- Data Magnitude: **MILLION** level
- Data Type & Storage

# Initial EER



## Final - EER



**Business:** stores restaurants business information and attributes



**Check-in:** stores times of customer check-in for each restaurant



**Yelp\_review:** store reviews generated from yelp customers



**Yelp\_users:** store user information of yelp customers



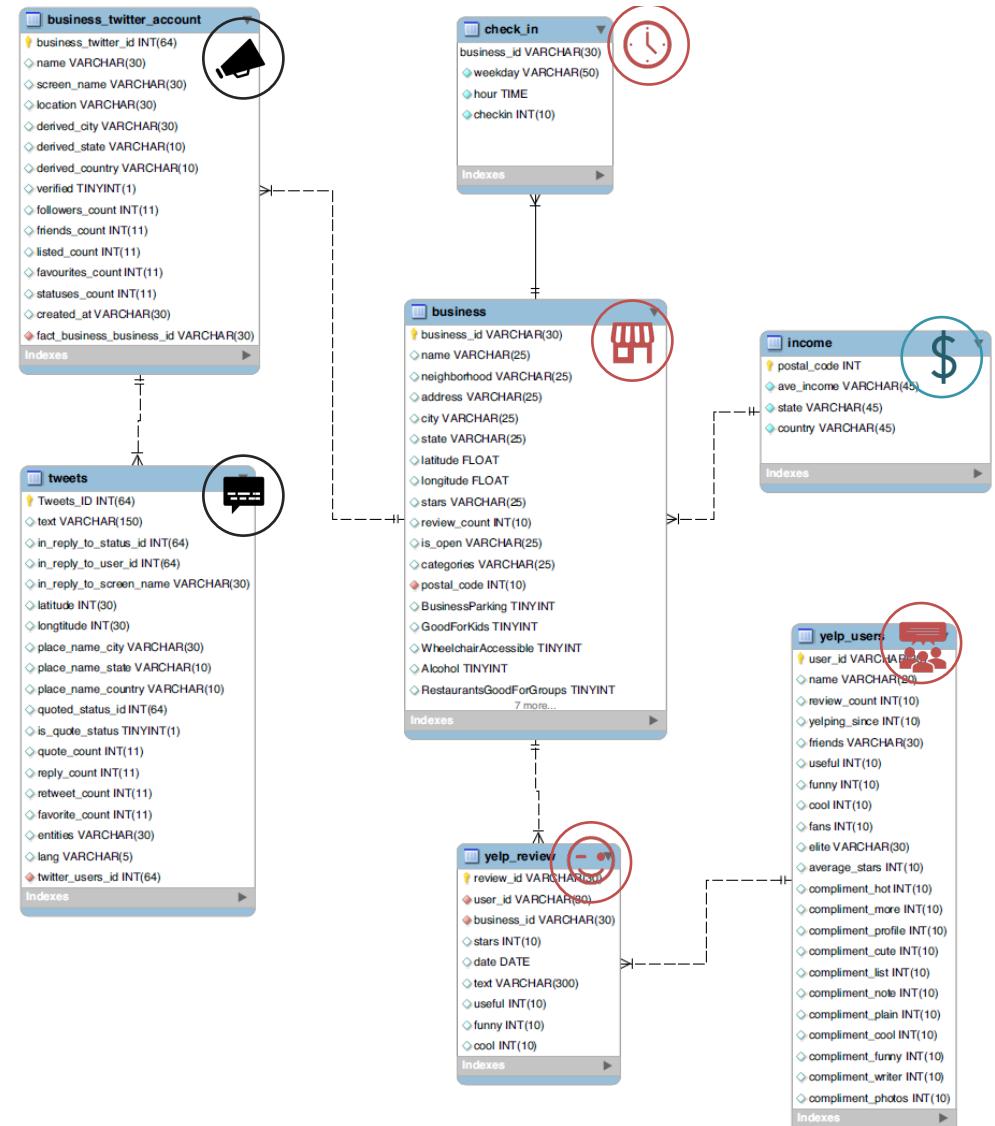
**Income:** stores income information for districts that restaurants are in



**Tweets:** stores tweets information and public impact



**Business\_twitter\_account:** stores restaurants information from its twitter account



## **Yelp Restaurants Industry Analysis**



How are restaurants geo-location, household income and Yelp review related?



Do usually low rating or high rating restaurants share similar categories?



What are the some most important restaurant attributes and how they impact Yelp ratings ?



Will Yelp review sentiment analysis result resonate Yelp rating?



Let's compare with Twitter!

## Geo-Locations and Yelp Ratings

Average rating of those states is around 3.5 stars.

Restaurants in higher income area may receive stringent reviews from customers

-> Location-targeting business expansion or franchising strategy



## Household Income and Yelp Reviews

Foodies from higher income areas are more willing to share food experiences on Yelp.

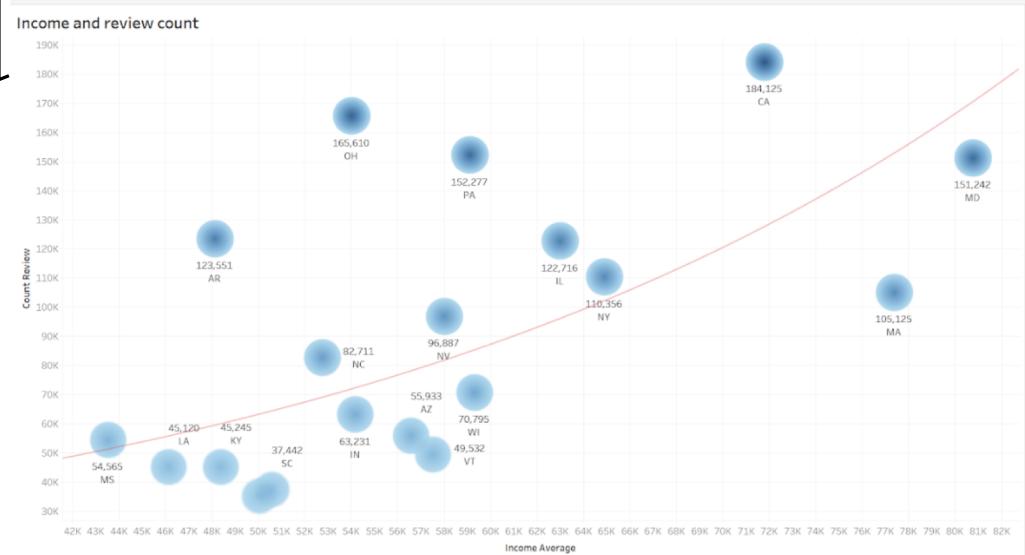
-> Public exposure and reputation management

**SELECT**

```
i.state,
AVG(stars),
SUM(review_count),
AVG(i.AverageIncome),
SUM(checkin)
```

**FROM**

```
business br
    LEFT JOIN
income i ON br.postal_code = i.zipcode
    LEFT JOIN
check_in c ON br.business_id = c.business_id
GROUP BY i.state;
```



STATE	postal_c...	AVG(cast(stars as sig...)	SUM(review_co...)	AVG(i.AverageInc...	SUM(chec...
AZ	85054	4.0000	1440	61121.66667000001	73
AZ	85202	3.0000	4914	144637.666699999...	275
AZ	85260	3.0000	7238	494079.8332999999	295
AZ	85281	2.0000	842	164020.83330000...	25
AZ	85345	3.0000	4067	190088.83330000...	72
IL	61866	2.0000	36	40802	23
NC	28078	4.0000	2115	471189.666700004	153
NC	28203	4.0000	26628	213039.833299999	672
NC	28213	2.0000	6	128946.3333	2
NC	28269	3.0000	708	356787.66670000...	93
NV	89052	3.0000	14904	603352.1667	490
NV	89109	4.0000	200238	206435.66670000...	2568
OH	44056	2.0000	56	80761.83333000002	18

## Categories that Most \* Restaurants Share

Focus on convenience, fast food or food that is filling but lack of keywords like food quality, service, dining experience



## Categories that Most ★★★★ Restaurants Share

Focus on dining experience, services and whether it is good for social environment.

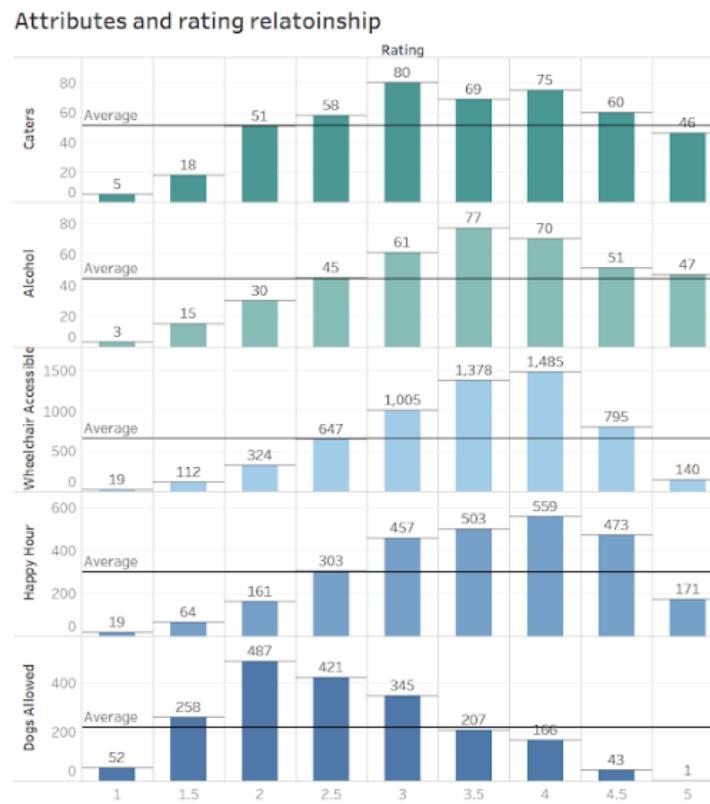
## Rating 5 wordcloud



## Attributes that Matters to the Yelp Ratings

Model comparison and variable selection using R.  
 -> More than 50 attributes narrowed down to top 5

Dogs Not Welcome?  
 Dining experience ↓



```

SELECT
    r.state, r.city, r.stars,
    CASE
        WHEN b.Alcohol = "True" THEN 1
        ELSE 0
    END Alcohol,
    CASE
        WHEN b.Caters = "True" THEN 1
        ELSE 0
    END Caters,
    CASE
        WHEN b.HappyHour = "True" THEN 1
        ELSE 0
    END HappyHour,
    CASE
        WHEN b.WheelchairAccessible = "True" THEN 1
        ELSE 0
    END Wheelchairaccessible,
    CASE
        WHEN b.DogsAllowed = "True" THEN 1
        ELSE 0
    END Dogsallowed,
    COUNT(*) restaurant_count
FROM
    attributes b
    LEFT JOIN
    business r ON b.business_id = r.business_id
group by state, city, stars;

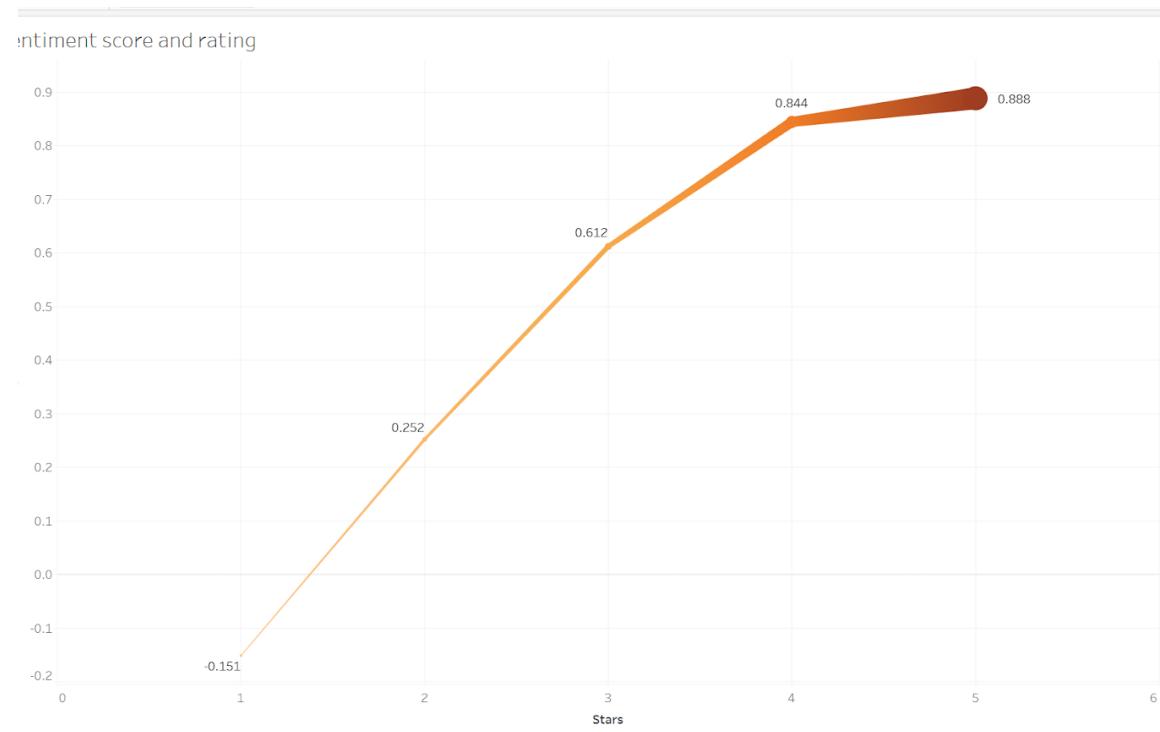
```

state	city	stars	Alco...	Caters	HappyH...	Wheelchairacces...	Dogsallo...	restaurant_c...
AZ	Ahwat...	4.0	0	0	0	0	0	3
AZ	Anthem	1.5	0	0	0	1	0	3
AZ	Anthem	2.0	0	0	0	0	0	2
AZ	Anthem	2.5	0	0	0	0	1	2
AZ	Anthem	3.0	0	0	0	0	0	1
AZ	Anthem	3.5	0	0	0	0	0	3
AZ	Anthem	4.0	0	0	0	0	0	2
AZ	Anthem	5.0	0	0	1	0	0	1
AZ	Avond...	1.0	0	0	0	0	1	1
AZ	Avond...	1.5	0	0	0	1	0	5
AZ	Avond...	2.0	0	0	0	0	0	16
AZ	Avond...	2.5	0	0	0	0	0	29
AZ	Avond...	3.0	0	0	0	0	0	27

## Sentiment Scores and Yelp Ratings Resonate

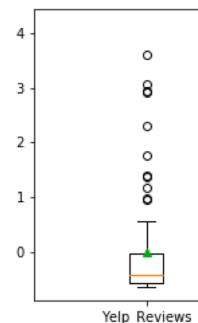
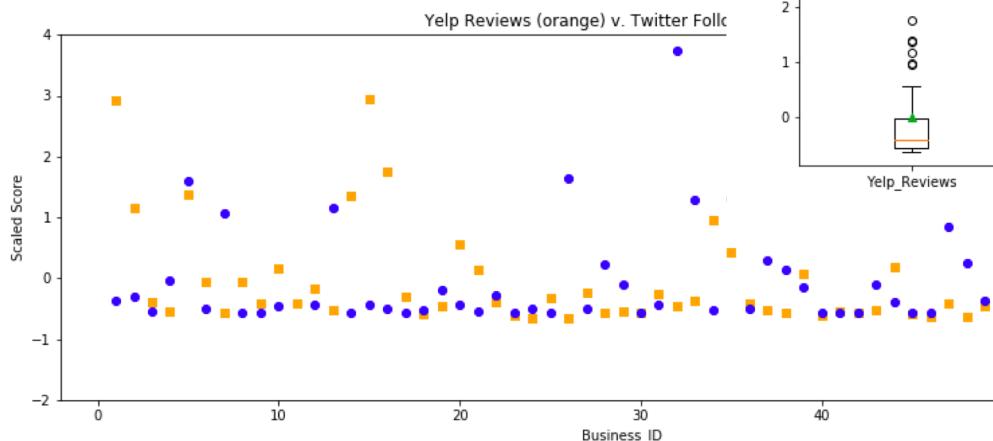
Not surprise!

Higher rating restaurants  
usually have higher score  
from sentiment analysis.



Yelp and Twitter share the similar pattern of Popularity

- Comparison between Yelp reviews\_count and Twitter followers\_count
- Both data are scaled
- 40% pairs have the gap less than 0.2 std,  
while 60% pairs have the gap less than 0.5 std



```

1 • SELECT
2     b.business_id,
3     t.twitter_account,
4     b.name,
5     b.stars,
6     COUNT(DISTINCT re.review_id) AS yelp_reviews,
7     t.followers_count AS twitter_followers
8
9     FROM
10    business AS b
11   INNER JOIN
12    yelp_review AS re
13   INNER JOIN
14    business_twitter_account AS t
15   ON
16    b.business_id = re.business_id AND
17    b.business_id = t.business_id
18   GROUP BY
19    b.business_id, t.twitter_account;
  
```

business_id	twitter_account	name	stars	yelp_reviews	twitter_followers
-2Arz8twKJmxHMS3S...	@SteepNBrew...	Steep & Brew West	3.5	4	112
-49WY_TEa9ZEcRk...	@CrackerBarrel	Cracker Barrel Old...	3.5	8	119240
-5L8zOxibac-vBrsYtxX...	@edstavernclt	Ed's Tavern	3.5	22	1898
-73xAVIIMh_O2nEAcG...	@Hardees	Hardee's	3.0	1	74153
-7d3UgQYYcBxbDH2d...	@noodlescom...	Noodles & Company	3.5	21	57558
-8Oaw94x2pYJCVP8Z...	@Carrabbas	Carrabba's Italian Grill	3.5	20	24750
-9na128tnoylwViuJVYEQ	@TapasPapafrita	Tapas Papa Frita	3.5	40	560
-AD5PiulHgdUcAK-Vx...	@Smashburger	Smashburger	3.5	21	25003
-Bb3kfdHX1qm2RvIH...	@jimmyjohns	Jimmy John's	3.0	1	488624
-bKnO1BwkwXjFNLr7...	@Brueggers	Bruegger's Bagels	3.5	4	11649
-B_4phIQkzU12AMaY...	@OCBBQ	Old Carolina Barbe...	3.5	7	839
-cBKodqj77Q0vk-9ID...	@fat_tues_day	Fat Tuesday	3.5	32	5763
-cKjQJsU73tlLKFA4un...	@Qdoba_Fres...	QDOBA Mexican Eats	3.0	4	157
-dGORHijPrnQzTZ2...	@GreatWraps	Great Wraps	3.0	4	1787
-DkhPijabXuTNvNtel...	@CommonGro...	Common Grounds	3.5	8	9614

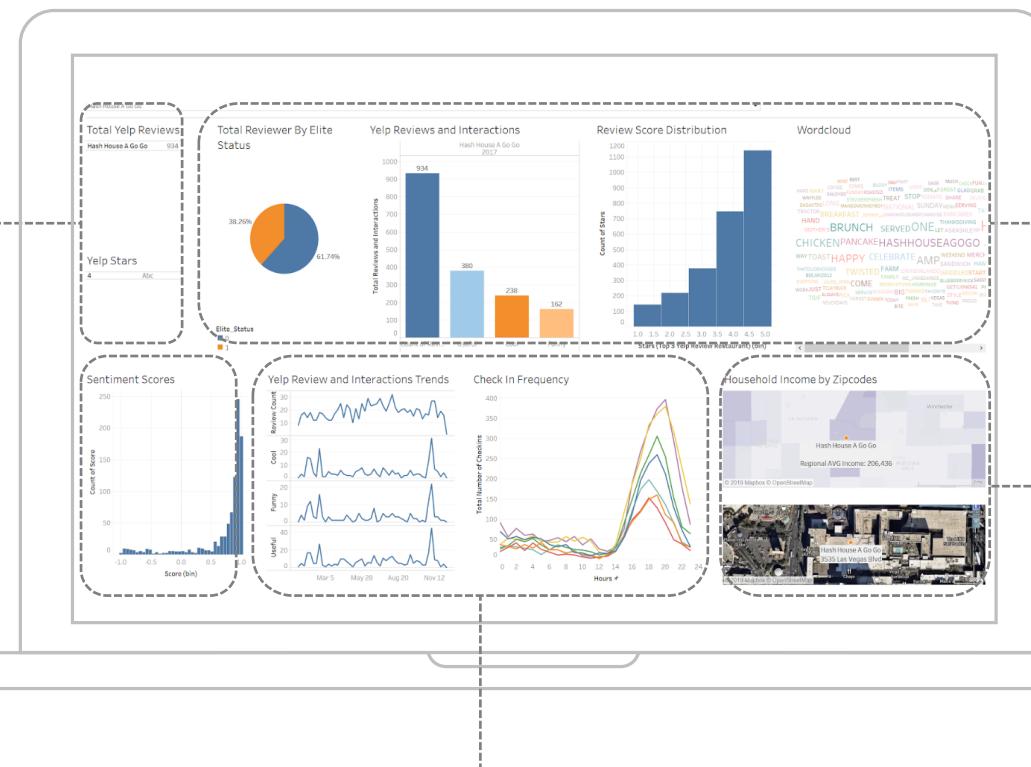


## Sentiment Monitoring

PR management for user influence, reviews sentiment and social media interactions

## Scoreboard

Summary of overall ratings and reviews

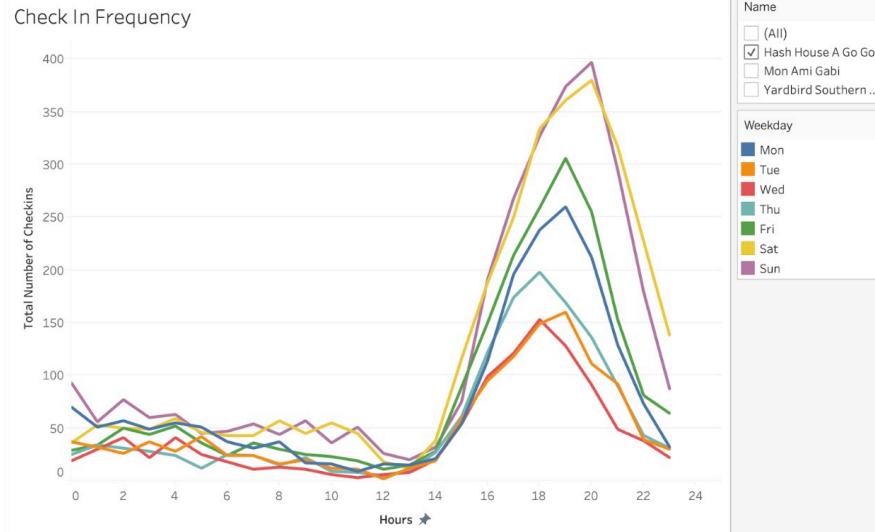


## Trending Tracker

Define trends and peak for timely marketing promotions

## Regional Demographics

Third party Community information and impact on the community



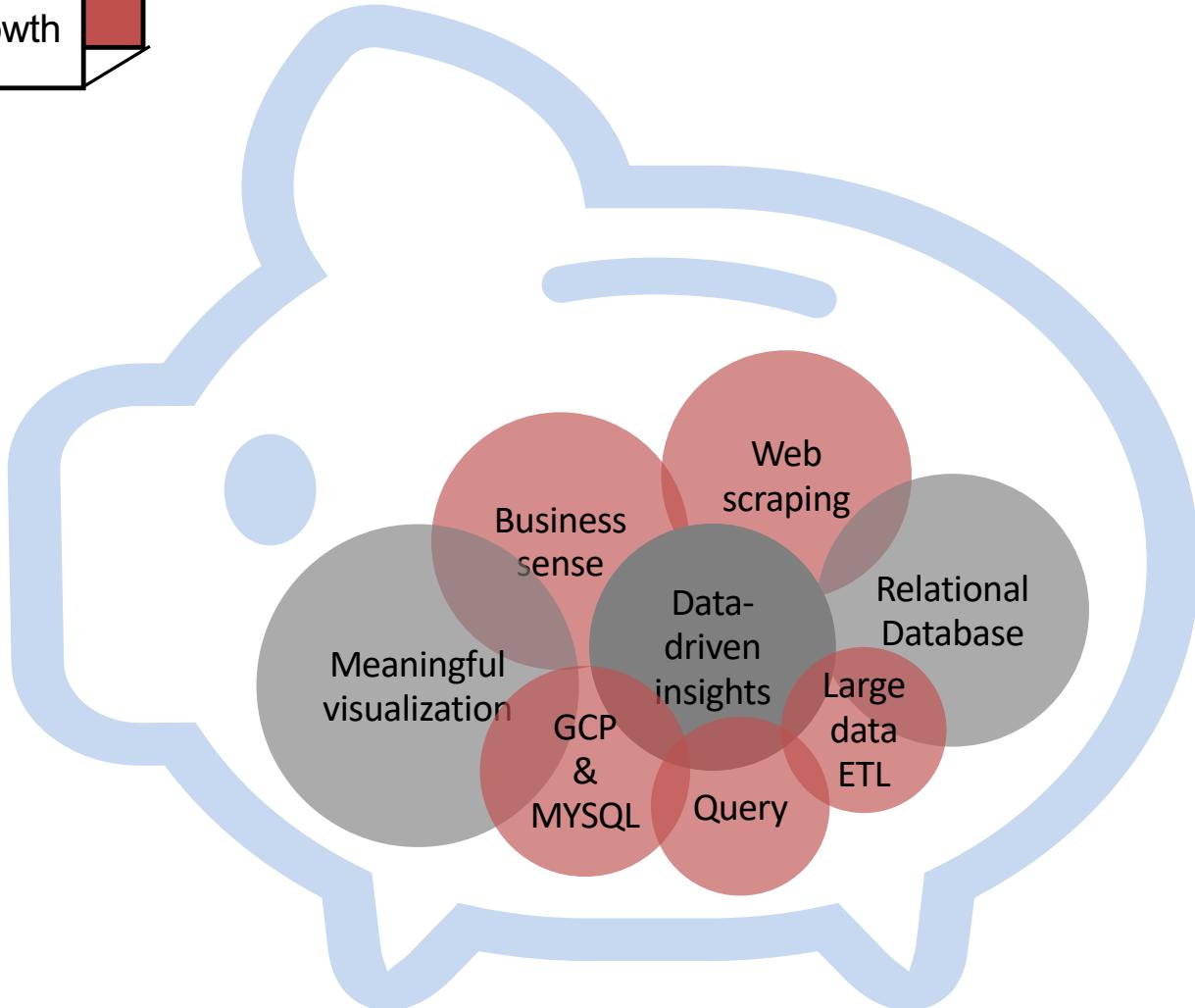
```

SELECT
    b.business_id,b.name,b.city, c.hour,
    CASE
        WHEN c.weekday = 'Sun' THEN SUM(c.checkin)
    END Sun,
    CASE
        WHEN c.weekday = 'Mon' THEN SUM(c.checkin)
    END Mon,
    CASE
        WHEN c.weekday = 'Tue' THEN SUM(c.checkin)
    END Tue,
    CASE
        WHEN c.weekday = 'Wed' THEN SUM(c.checkin)
    END Wed,
    CASE
        WHEN c.weekday = 'Thu' THEN SUM(c.checkin)
    END Thu,
    CASE
        WHEN c.weekday = 'Fri' THEN SUM(c.checkin)
    END Fri,
    CASE
        WHEN c.weekday = 'Sun' THEN SUM(c.checkin)
    END Sun
FROM
    check_in c
    LEFT JOIN
        business b ON b.business_id = c.business_id
GROUP BY b.business_id,b.name,b.city,c.hour,weekday;

```

business_id	name	city	hour	Sun	Mon	Tue	Wed	Thu	Fri	Sat
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	00:00:00	NULL	NULL	NULL	NULL	NULL	1	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	00:00:00	NULL	1	NULL	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	00:00:00	NULL	NULL	NULL	NULL	NULL	5	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	00:00:00	7	NULL	NULL	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	00:00:00	NULL	NULL	1	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	01:00:00	NULL	NULL	NULL	NULL	NULL	2	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	01:00:00	NULL	NULL	NULL	NULL	NULL	6	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	01:00:00	2	NULL	NULL	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	01:00:00	NULL	NULL	NULL	1	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	01:00:00	NULL	NULL	1	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	02:00:00	NULL	NULL	NULL	NULL	NULL	1	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	02:00:00	1	NULL	NULL	NULL	NULL	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	04:00:00	NULL	NULL	NULL	NULL	NULL	1	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	17:00:00	NULL	NULL	NULL	NULL	1	NULL	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	17:00:00	NULL	NULL	NULL	NULL	NULL	4	NULL
-7zmmkVg-IMGaXbu...	Primal Brew...	Huntersv...	17:00:00	5	NULL	NULL	NULL	NULL	NULL	NULL

## Knowledge Bank Growth



## Future Consideration



### NOSQL DATABASE

Text data, such as reviews or tweets, is more suitable to store in the nosql system



### SCALING WITH GCP WORKFLOW INTEGRATION

Create auto-workflow rely on GCP apps, scale up the applications to a bigger market



### ENRICH DATA SOURCES

Google Review  
Facebook  
Instagram  
CRM



### ENHANCED ANALYSIS AND TARGETING

NLP  
Persona  
People-based targeting

