



# FROOSN

Identifying Fake Reviews on OSNs(Yelp)



## **Contributors**

Sahil Babbar

Akash Kumar Gautam

Mayank Kumar

Shyam Agrawal



# Motivation

behind the project!





# Motivation



BUSINESS

## Yelp cracking down on businesses that reward reviewers

## Fake Yelp reviews: anatomy of an 'astroturfing' post

It's no surprise that people are paid to write fake internet reviews - but how do you spot astroturfing? Here's our guide to help

Incentivised reviews are warping Amazon's product star ratings, report says

Reviewers given discounts or free products award higher ratings, detrimentally affecting Amazon's star system

JUN 27, 2016 @ 04:44 PM 2,580

## The Danger Of Incentivizing Customer Reviews



Steve Olenski, CONTRIBUTOR

I write about advertising, marketing, media & all subgroups therein. [FULL BIO](#)

Opinions expressed by Forbes Contributors are their own.

## Amazon bans 'incentivised' reviews over lack of impartiality

Evidence shows reviewers awarded higher star ratings to products that were discounted in exchange for a review



# Motivation

Support Center

## Don't Ask for Reviews

You may be hurting your Yelp rating by proactively asking your customers for reviews.

 **ReviewTrackers**  
We're hiring!

Product

Plans

Industries

Resources

## Requesting Customers to Review Your Business? Don't Incentivize.

By Migs Bassig on March 7, 2017

**HUFFPOST**

**THE BLOG** 05/27/2016 11:55 am ET | Updated May 28, 2017

## Are Online Reviews Reliable? (Or Understanding Astroturfing)



Where  
do we intend to work?





## Yelp Open Dataset

An all-purpose dataset for learning



### The Dataset



4,700,000 reviews



156,000 businesses



200,000 pictures



12 metropolitan areas



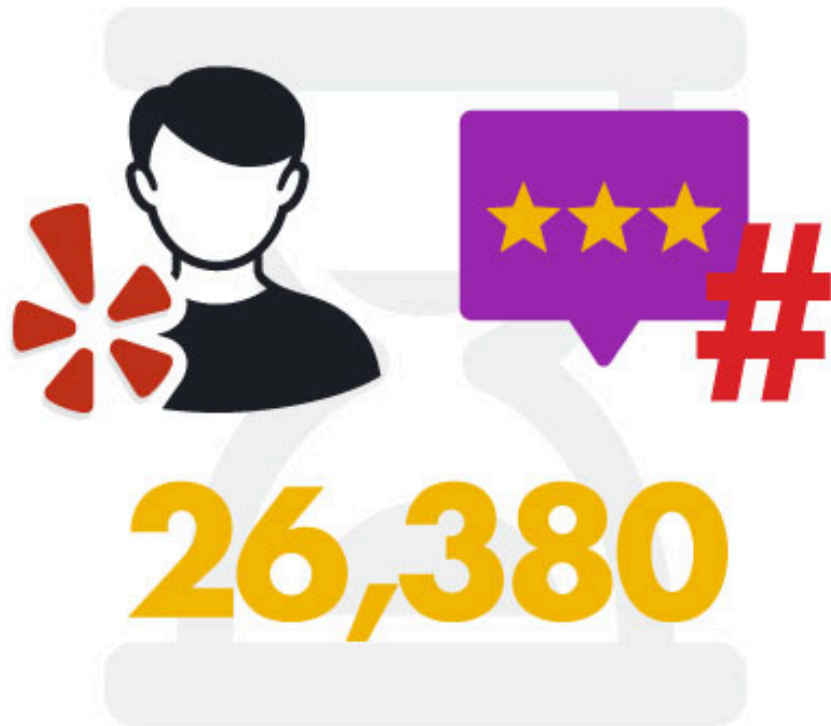
# Why

are we doing, what we are doing?

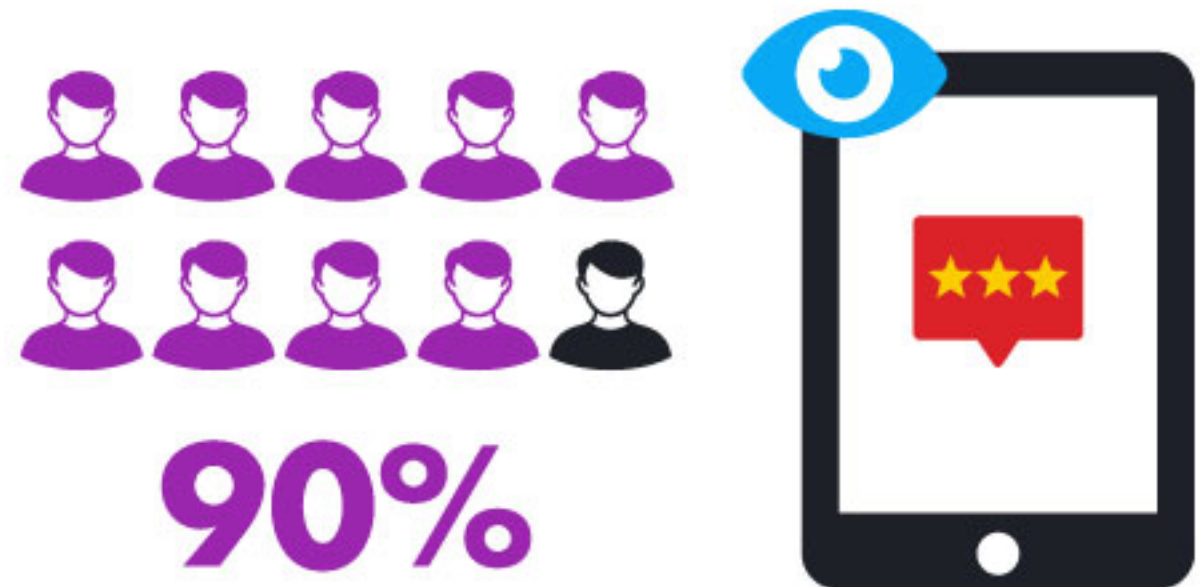


# Why?

Number of reviews posted every minute by Yelp users - **26,380**



**90%** of consumers read online reviews before visiting a business.



**86%** of people will hesitate to purchase from a business that has negative online reviews.



Compiled & designed by:





**88%** Of Consumers Trust Online Reviews As Much As Personal Recommendations.



**88%**

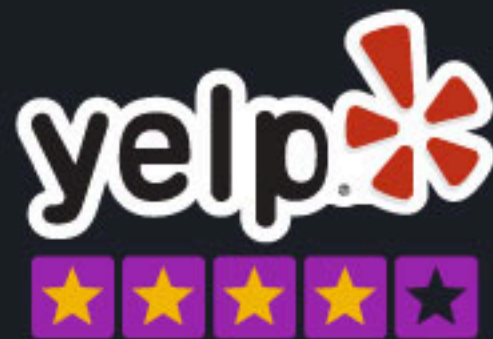
**72%** say that positive reviews make them trust a local business more.



**72%**



On average, a one-star increase on Yelp leads to a **5 to 9** percent increase in a business's revenue, one negative review can cost you **30 customers**.



**one** negative review

=



can cost you **30 customers**



# What

are we doing & what is the scope?





# What?

---

For this project

Fake = Incentivised

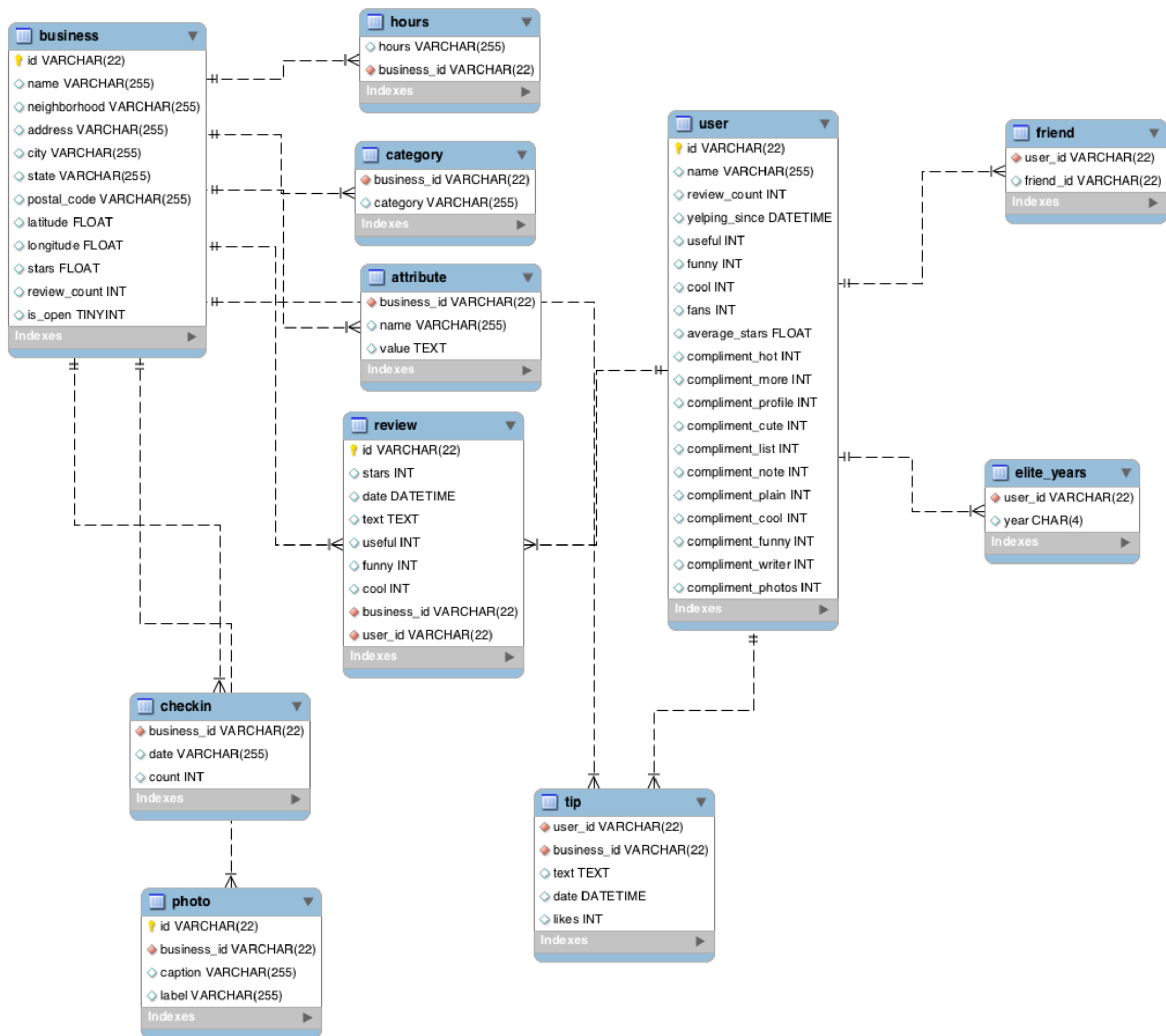
OSN = Yelp

## Problem Statement

“Building a mechanism to identify fake reviews among the set of all user reviews, in order to maintain the authenticity of online businesses.”

## Scope

- We assume that **there are no bots**, so everyone is a legit user.
  - Yelp has very stringent security provisions, for e.g. it detects anonymous network browsing, web scraping, tracks IP addresses, does automated filtering, etc.
- We can only **leverage the attributes of data** provided in the Yelp Dataset and Yelp Business API and anything outside those, cannot be analysed.





# How

are we achieving our end goals?





# How?

---

## Methodology

**Step 1:** URL of the business/review is input into the system.

`https://www.yelp.com/biz/gary-danko-san-francisco`

**Step 2:** Parse the URL and find the business ID.

`'leJWawvw4oAb9EjAP-W5hA'`

**Step 3:** Calculate the following 4 metrics for the business:

User Rating Deviation

Business Rating Trend

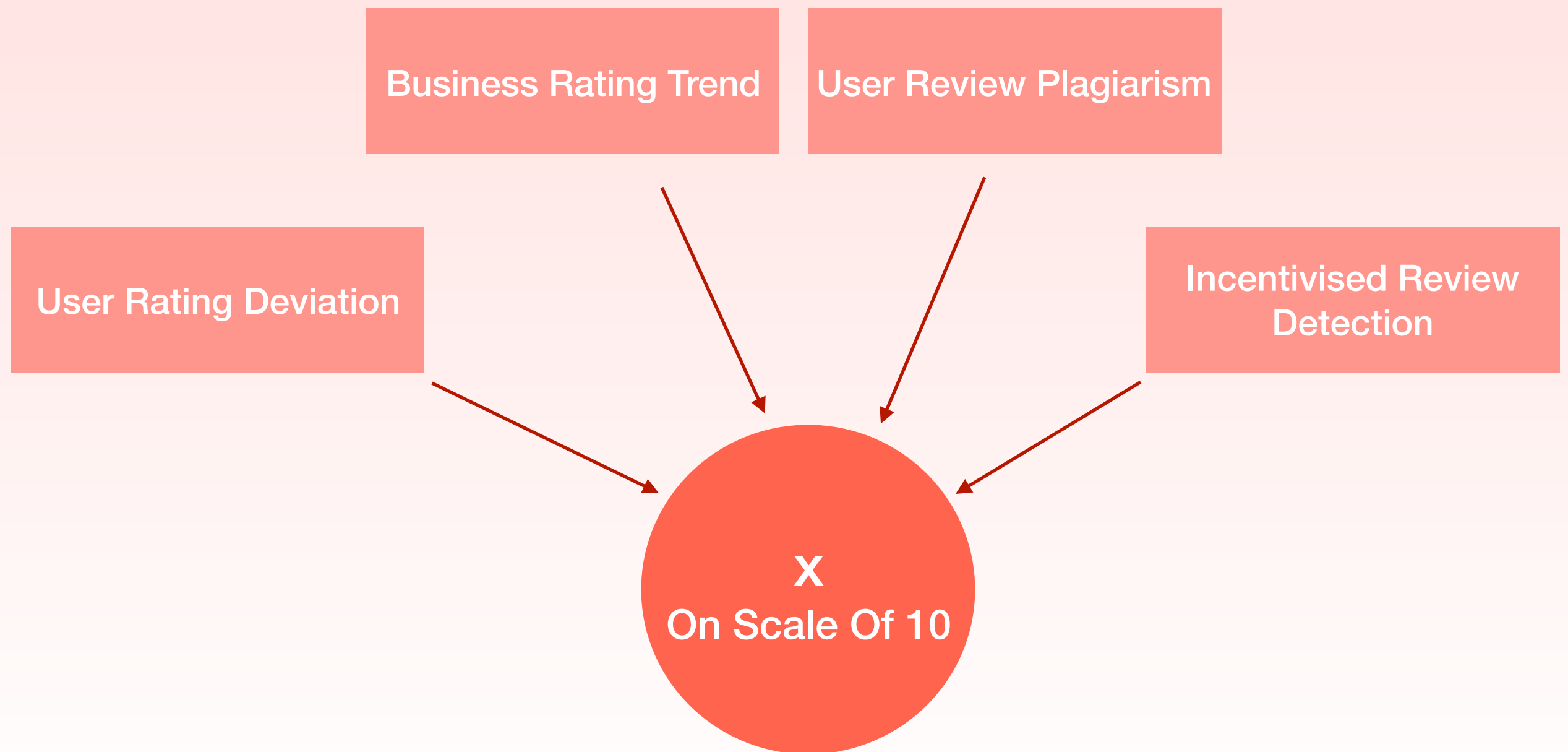
User Review Plagiarism

Incentivised Review  
Detection





**Step 4:** Calculate the credibility score for all the reviews present on the business's profile page.







**Step 5:** Show the visualisations in the form of graphs and statistics of the analysis done in Step 3.



# Initial Work

done till now!





# Preliminary Work

---

**Data Collected**



# Preliminary Work

---

**Analysis Done**



# Preliminary Work

---

## Examples



# Next

things on our to-do list!





# Next Steps

---

- Calculating the remaining metrics as mentioned earlier.
- Real-time review analysis. So that businesses not in Yelp Data Set, can also be analysed.
- Use phrase-repetition detection, in plagiarism analysis.



# Your Feedback

is indispensable for us!

