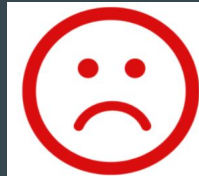


# Sentiment Analysis of Insurance Apps Reviews

- Data Science Intensive Capstone



Clear and crisp instructions are given. Easy to claim. Very much compact but informative. Quick approval obtained each time. App is too good to work with.



Waste application I have seen ever. We can't renew our policy also app is dead slow every time.

By- Rohan Prabhakar Agale

Mentor - Ben Bell

Reviewer - Andrea Constantino

# Executive Summary

## Problem Statement

Mobile applications play a crucial role in enhancing customer experience and offering customized products. Using Play store reviews of Indian GI companies' app, we analyze what is working, what are the challenges and how companies can improve the apps.

## Insights

Many customers are facing two basic but major challenges.

1. App is slow / does not open.
2. Could not add / view policy details

Customers are happy and vocal about in app tele-consultation services by doctors & dieticians.

## Business Actions

With rigorous worst case scenario testing and better backward integration with IT systems, companies can provide satisfactory user experience.

By increasing mobile app adoption & providing teleconsultation like services, companies can provide enhanced experience to the customers.

## Data

We scraped mobile app reviews of Indian GI (General Insurance) companies from Google Play Store.

## Modelling

MultinomialNB with count vectorizer beats XgBoost and ties up with Logistic regression. In MultinomialNB, impact of predictive features is directly interpretable. Hence, it is the best model for our purpose. It generalizes well with **pr\_auc** of **0.944** on test data.

# Objective

## Context

Number of smartphones users in India grew to 438M in 2020 with a healthy YoY growth rate of 27%. Almost 80% of screen time on smartphones is spent on using different mobile apps. Hence, General Insurance companies have build mobile apps which offer convenient and seamless channel to buy insurance policies and intimate claims.

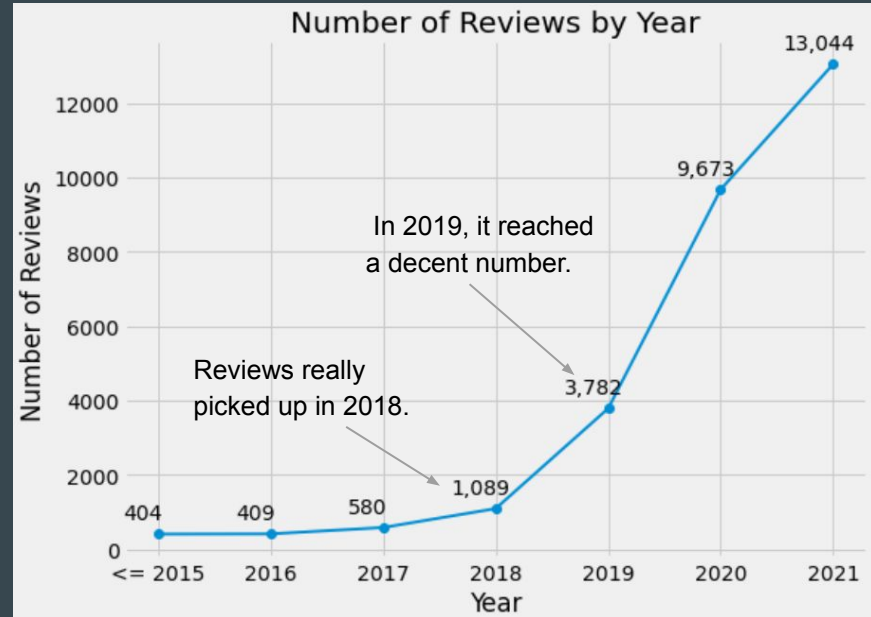
## Problem Statement

Leverage Play Store reviews to analyze customers' feedback to know

1. what is working,
2. what are the challenges and
3. how companies can address those challenges.

## Why

With mobile apps, companies can offer instant services, increase customer engagement, build loyalty and offer customized products. Hence, it is important for the company to know how well the app is achieving the objectives.



\*Top 5 GI companies & 3 InsureTechs (93% reviews)

\*\* 2021 is as of 26 Aug 2021.

# Data Scraped from Play Store

- Identify the company using a unique id.

<https://play.google.com/store/apps/details?id=com.ba.cp.controller&showAllReviews=true>

- Using `google_play_scraper` api, we will one by one extract reviews for all the companies. Data sample is below

| reviewId                            | content   | score | thumbsUpCount | reviewCreatedVersion | at                     | replyContent | repliedAt | appld             |
|-------------------------------------|---|-------|---------------|----------------------|------------------------|--------------|-----------|-------------------|
| YINfO2IN5pk3baCjf8...               | Excellent..<br>The service<br>and<br>reaction<br>were exce... | 5     | 0             | NaN                  | 2021-08-26<br>22:29:04 | NaN          | NaN       | icici.lombard.ghi |
| vHSV2KRUCILU1eSsx-5HzWABJDC2pCW6... | The app is<br>simple and<br>tidy, with a<br>focus on<br>th... | 5     | 0             | NaN                  | 2021-08-26<br>22:28:35 | NaN          | NaN       | icici.lombard.ghi |
| zuDrtrQcYBBJSkmZ...                 | It is simple<br>to select<br>and<br>purchase<br>insurance.... | 5     | 0             | NaN                  | 2021-08-26<br>22:27:21 | NaN          | NaN       | icici.lombard.ghi |

- For our analysis, we use **only English language reviews**.

Data Wrangling

EDA

Modelling

Actionable Insights

# Company Ratings :



For the lack of sufficient & credible reviews, we consider top 3 private players (ICICI Lombard, Bajaj Allianz & HDFC Ergo) and one insurtech (Acko) reviews after 2018.

Acko is the leader with 3.8 stars.

ICICI Lombard & Bajaj Allianz score ~3.5 stars. These GI companies are adopting & making good progress in the mobile application space.

HDFC ERGO is at 2.8 stars & needs to catch up with the leader. It has highest reviews of 7097.

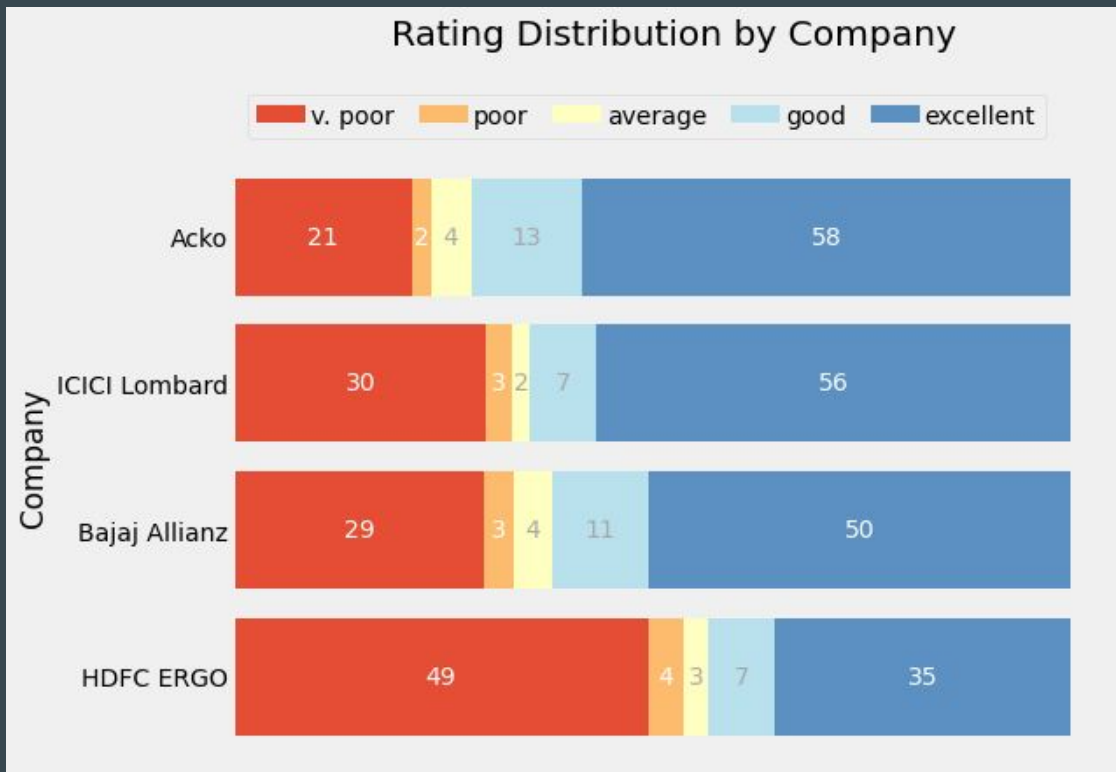
Data Wrangling

EDA

Modelling

Actionables

# Review Distribution is Polarized



Acko & ICICI Lombard have 58% & 56% 5 star ratings. Yet these companies have a 21% & 30% 1 star rating for apps. This indicates a huge scope of improvement even for leaders.

HDFC Ergo has 49% reviews with just 1 star rating.

We have a highly polarized dataset. 80-85% of the reviewers have given either Excellent or Very Poor rating. Only about 3% reviewers are neutral. Therefore we

1. combine good & excellent as label 0
2. combine poor & very poor label 1
3. ignore average (809 reviews)

Data Wrangling

EDA

Modelling

Actionables

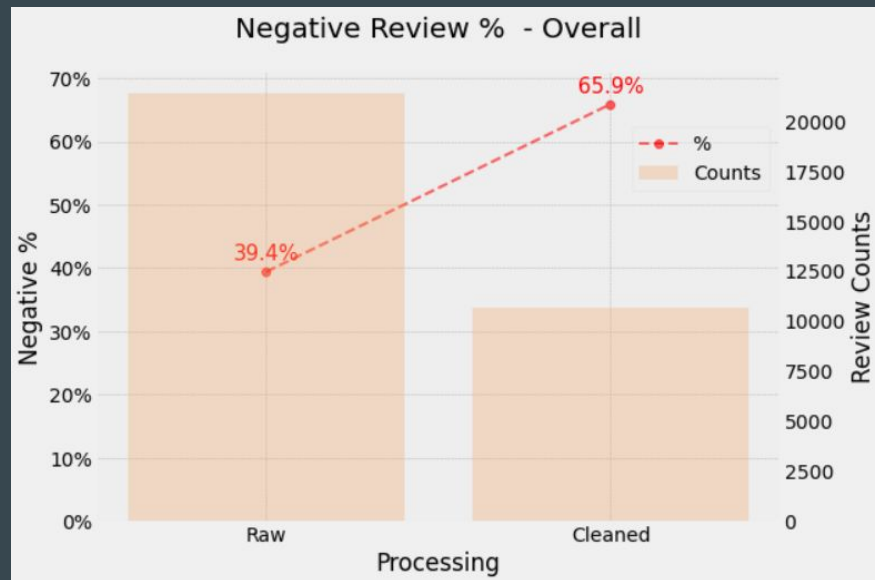
# Text Cleaning

## 1. Remove short reviews( < 4 words)

```
['Nice 🍑' 'nice app' 'good one' 'Very good' 'Very good']
```

2. Remove leading and trailing whitespaces
3. Expand contractions. Eg I've is converted to I have
4. Convert emojis to text
5. Convert emoticons to text
6. Convert text to lower cases
7. Remove non-alphanumeric text. This will remove reviews in Indian languages
8. Remove brand names. Eg ICICI Lombard, Acko etc
9. Remove extra white spaces.
10. Lemmatize the text
11. Removed stop words
12. Removed reviews with < 3 words.

Reviews are reduced from 21370 to 10640. Only 634 negative reviews are removed. Makes sense as customers with positive experience would write short reviews with generic words.



% of Negative ratings increased by 66%. Our main objective is to find what is driving the negative sentiments. Hence, it is OK if the probability of negative reviews changes.

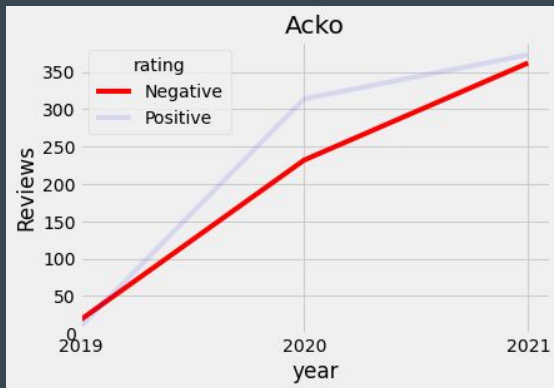
Data Wrangling

EDA

Modelling

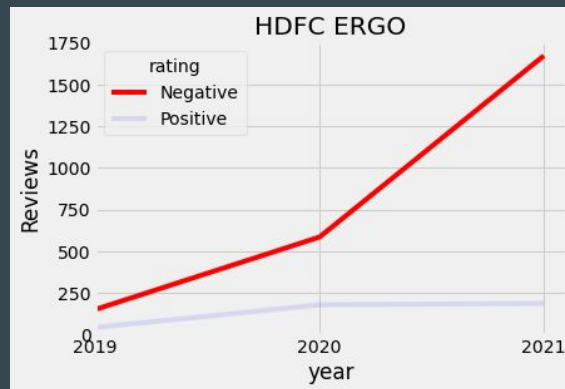
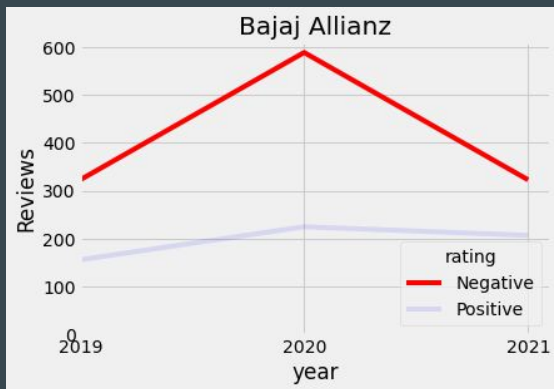
Actionables

# Company wise trends in negative reviews :



For Acko & ICICI Lombard, growth in negative reviews is in line with the trend.

Bajaj Allianz has significantly brought down negative reviews by half in 2021.



While HDFC Ergo has a clear problem of almost 3 times increase in negative reviews in 2021.

\*\*EDA is carried on train data(0.75).  
We stratified the data by label & company.

Data Wrangling

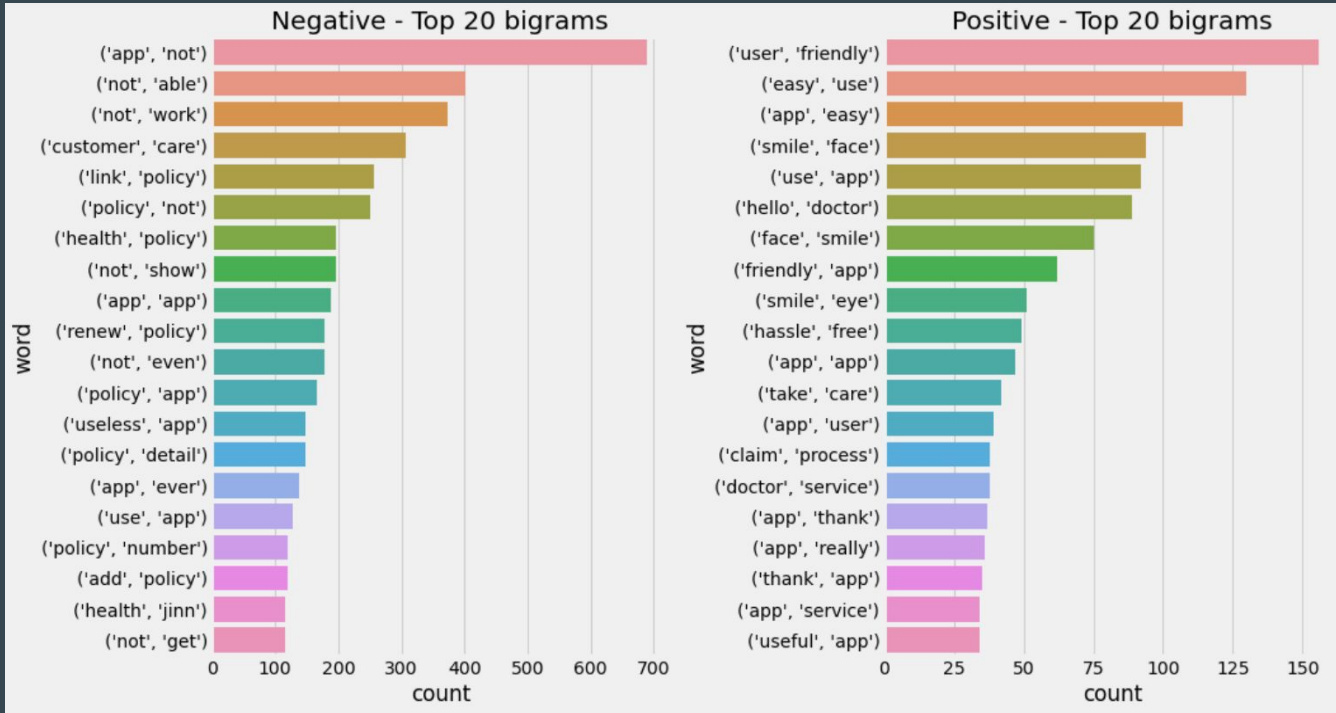
EDA

Modelling

Actionables



# Most Frequent Bigram by Sentiment :



Analysis based on top 15 words in 1-gram, bigram & trigram

## Negative Sentiment

1. App is not working as expected.
2. Unable to link / renew policy.
3. Login issues
4. App is slow.
5. Lacking Customer service.

## Positive Sentiment

1. Ease of use
2. Hello Doctor service
3. Claims process

Data Wrangling

EDA

Modelling

Actionables

# Models Comparison :

Our primary focus is understanding reasons behind negative sentiments. Therefore, we use the **pr\_auc matrix** as it gives more importance to positive class which in our case is negative sentiment.

| Best Model from RandomSearchCV ( 3 Fold CV) | Vectorizer<br>ngram range : (1,4) | Iterations | Mean pr_auc score | Standard deviation of pr_auc score |
|---|-----------------------------------|------------|-------------------|------------------------------------|
| Logistic Regression                         | Count (min_df=80)                 | 20         | 0.941243          | 0.003511                           |
| Logistic Regression                         | TfIdf (min_df=160)                | 20         | 0.938307          | 0.001265                           |
| Multinomial Naive Bayes                     | Count (min_df=80)                 | 20         | 0.940180          | 0.003642                           |
| XgBoost                                     | TfIdf (min_df=160)                | 20         | 0.938992          | 0.000500                           |

Logistic Regression with count vectorizer and MultinomialNB models perform well and are within 1 standard deviation of each other. We will choose **MultinomialNB** because it is **directly interpretable** and the contribution of each feature towards the negative class is very clear. Also, conditional probabilities are **easier to explain to business** people than LR coefficients. **0.94440 score on test data** is slightly greater than 0.94018 cross validation score but is about 1.2 std deviation(0.0036) from it.

Data Wrangling

EDA

Modelling

Actionables

# Feature Importance by Sentiments : Top 15

## Negative

| words         | P(Negative   word) |
|---------------|--------------------|
| slow app      | 0.999784           |
| loading       | 0.984085           |
| work properly | 0.966777           |
| exist         | 0.966385           |
| every time    | 0.957028           |
| show policy   | 0.950579           |
| add policy    | 0.947608           |
| load          | 0.945043           |
| login         | 0.943337           |
| nothing       | 0.940099           |
| open          | 0.936334           |
| ask           | 0.933147           |
| many time     | 0.931878           |
| never         | 0.929929           |
| properly      | 0.927458           |

## Positive

| words         | P(Negative   word) |
|---------------|--------------------|
| app easy      | 0.037520           |
| easy use      | 0.040844           |
| easy          | 0.046129           |
| quick         | 0.053958           |
| thank         | 0.055171           |
| doctor        | 0.058814           |
| helpful       | 0.116591           |
| simple        | 0.133616           |
| fast          | 0.139866           |
| happy         | 0.182548           |
| friendly      | 0.186852           |
| free          | 0.189617           |
| user friendly | 0.200088           |
| feature       | 0.225482           |
| user          | 0.327261           |

## Negative Sentiment

1. The app was **slow / not loading**.
2. The app was **not showing policy / claims** details.
3. Customers **could not add policy** in the app.
4. Some **could not login**.
5. The app was not working as per expectations.

## Positive Sentiment

1. Hello **doctor tele-consultation** service is helpful
2. The app was **easy / simple / friendly**.
3. The app was **fast / quick**.

\* Included company name as dummy variable to control for company effect.

\*\* Logistic Reg has almost same top important features

\*\*\* Separate modelling for HDFC Ergo & Acko leads to same conclusions

Data Wrangling

EDA

Modelling

Actionables

# Actionables

| # | Areas of Improvements                 | Solution   | Description  |
|---|---------------------------------------|--|--|
| 1 | App is slow or not loading or hangs   | Rigorous Worst Case Testing                      | <b>Make app work fast on all the devices.</b> Customers hate slow apps in this fast world. Company can use application logs to understand the issues like memory, internet speed etc. Accordings develop solutions and rigorously test them.             |
| 2 | Add / view / change policy or claims. | Better Backend Integration                       | <b>Make viewing and changing details available all the time.</b> Backend Server communication needs to be streamlined. Policy /claims details should be instantly available to the customers. Optimized offline/online data management can be exploited. |
| 3 | Offer Value added services            | Teleconsultation / Car Expert Advice             | <b>Provide useful Value Added Services.</b> Hello Doctor service by ICICI Lombard is well appreciated. Mobile apps provide instant access to these services.   |
| 4 | Increase Adoption of mobile apps      | Campaign to increase app installations and usage | <b>Make customers aware of the application.</b> Similar Teleclinic service by a competitor does not find any mention. With the awareness campaign about app and its features, companies can enhance customer experience.                                 |

Data Wrangling

EDA

Modelling

Actionables