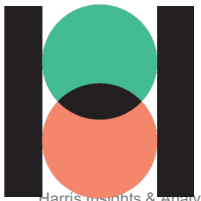
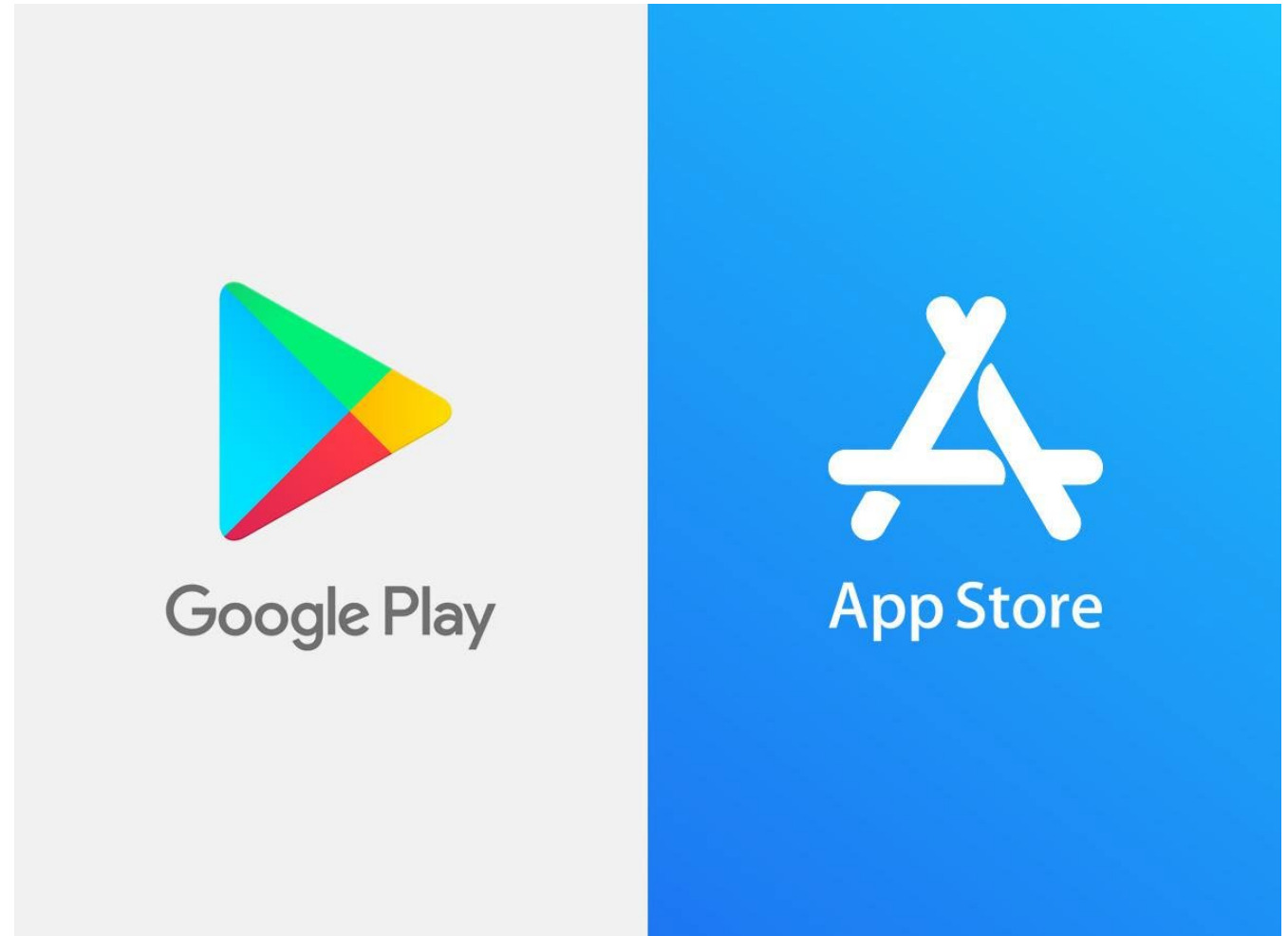




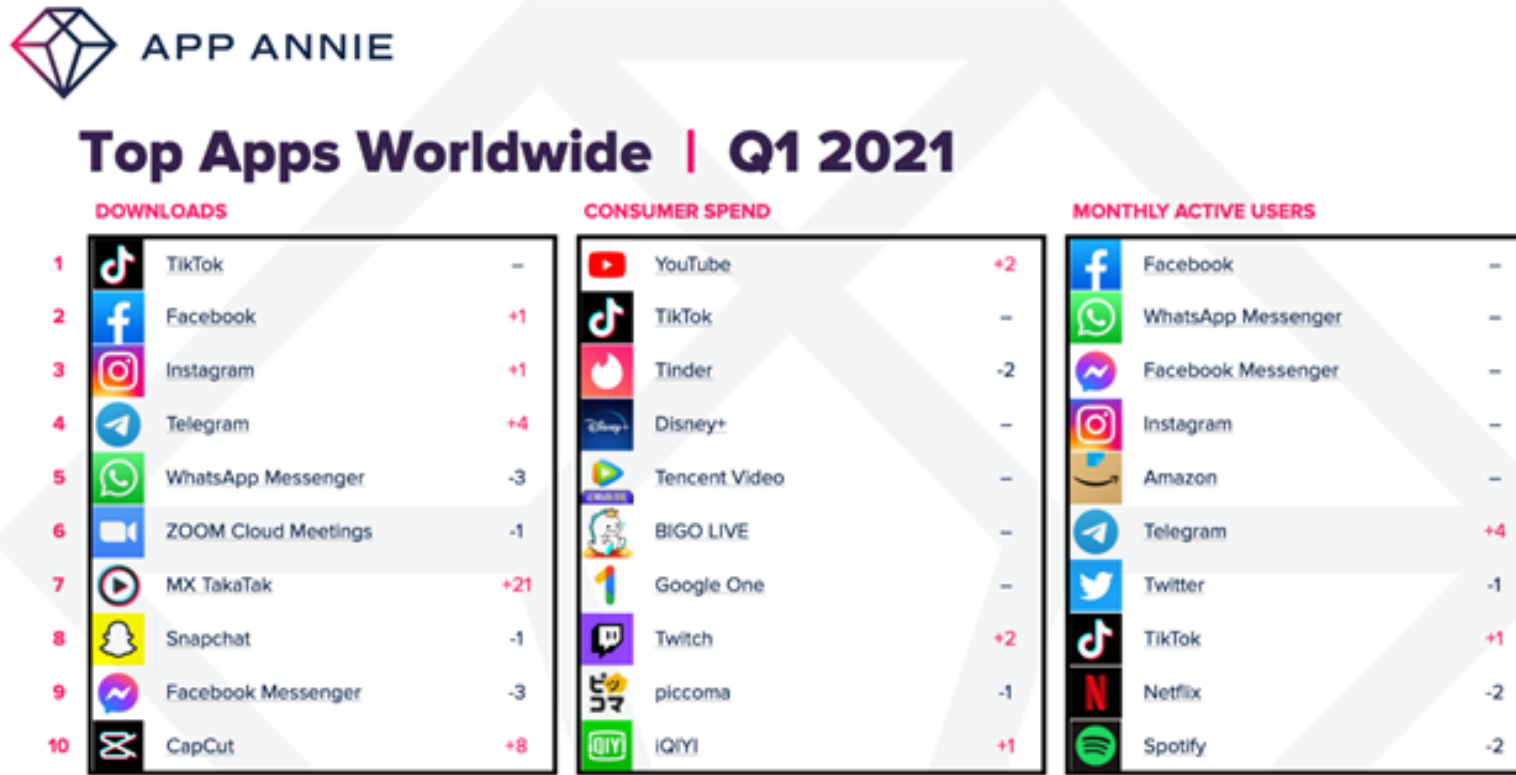
# Predicting App Adoption Rate Curves Using Time Series Modelling and Nearest Neighbors

Tomer Zur



# Introduction

- Obtaining a wide user base with frequent usage is so important
- More users → more advertising revenue and in-app purchases → more profitable app



*Note: Downloads and consumer spend based on combined iOS App Store and Google Play as of March 29<sup>th</sup> 2021 (iOS only for China). MAU based on iPhone and Android phone combined, excluding China, last full month of data (February 2021). Pre-installed apps (such as YouTube on Android phone and Safari on iPhone) are excluded from MAU ranking. All estimates from App Annie Intelligence.*

# Methods



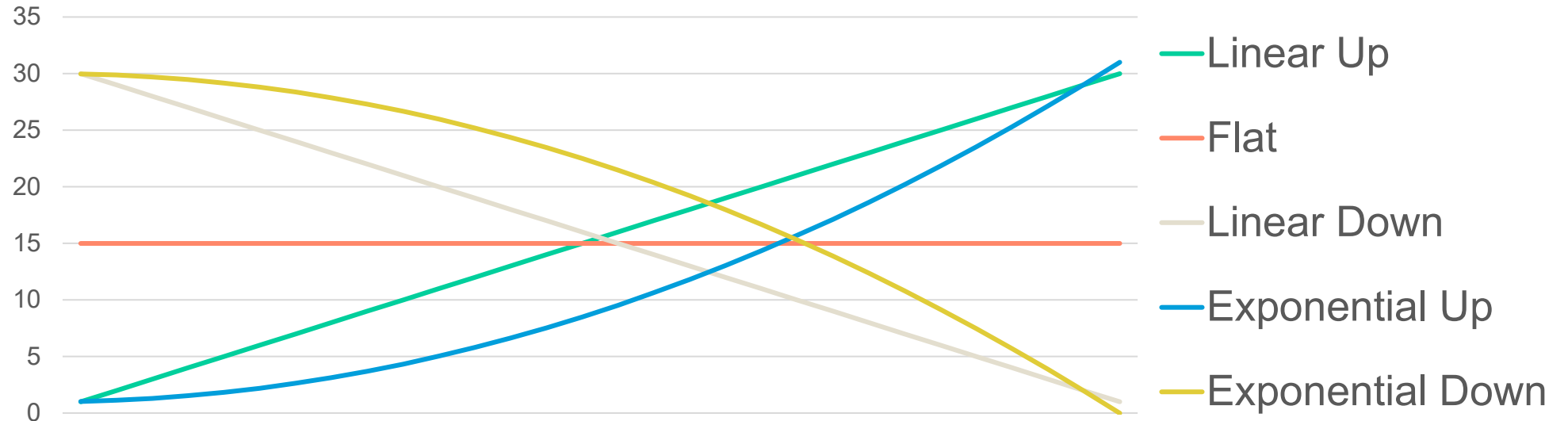
- My data - 10,000 permissioned respondents (from Verto) - usage tracked from Jan 2020 - Apr 2021
  - Dataset includes demographic info (gender, age, income, race, etc), app info (name, publisher, category), daily usage info by app (users, sessions, device split, downloads)

# Methods



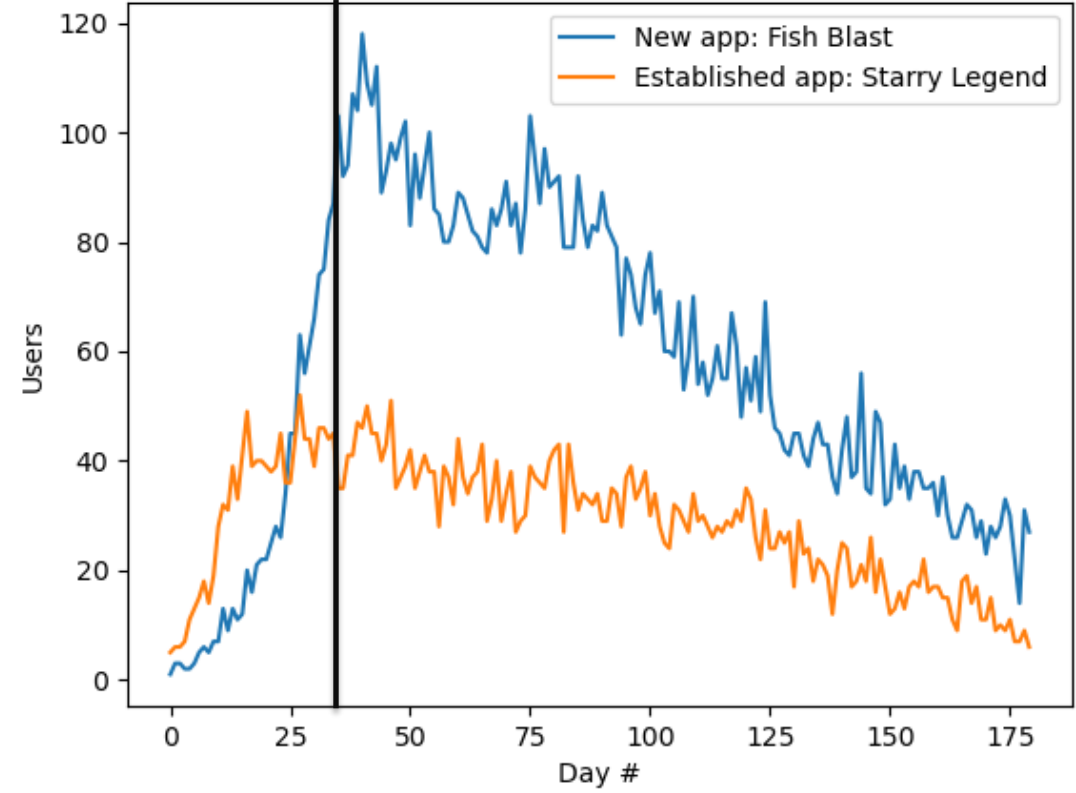
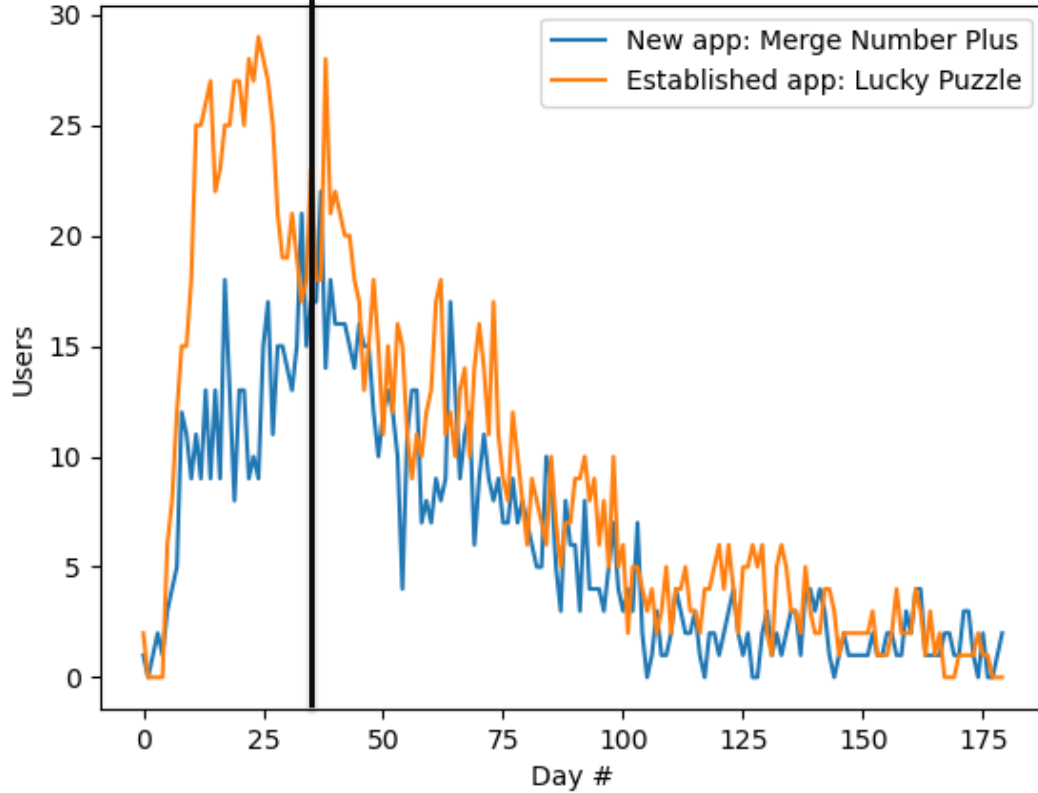
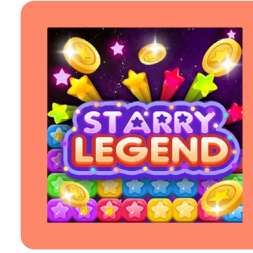
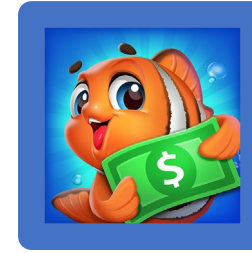
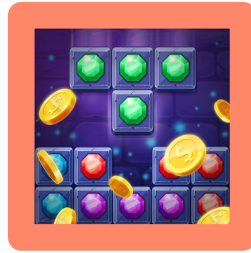
We will be:

- Clustering apps based on trend of users (exp. up, linear up, flat, linear down, exp. down)



- Fitting data (per app) to time series models (VAR, Polynomial Regression, Linear Regression, Facebook Prophet, Exponential, Ensemble) in each cluster
- Creating nearest neighbor model with each app's time series model coefficients as inputs
- Output will be prediction of users per day for next 150 days
- We will be measuring MSE, Accuracy, Precision, Recall of the models

# Sample Data



# Results – Using only apps released since 2020



- 5-fold Cross Validation used (since test dataset is very small n=20)
- When using all features (averages over 5 folds):



## Results – Using only apps released since 2020



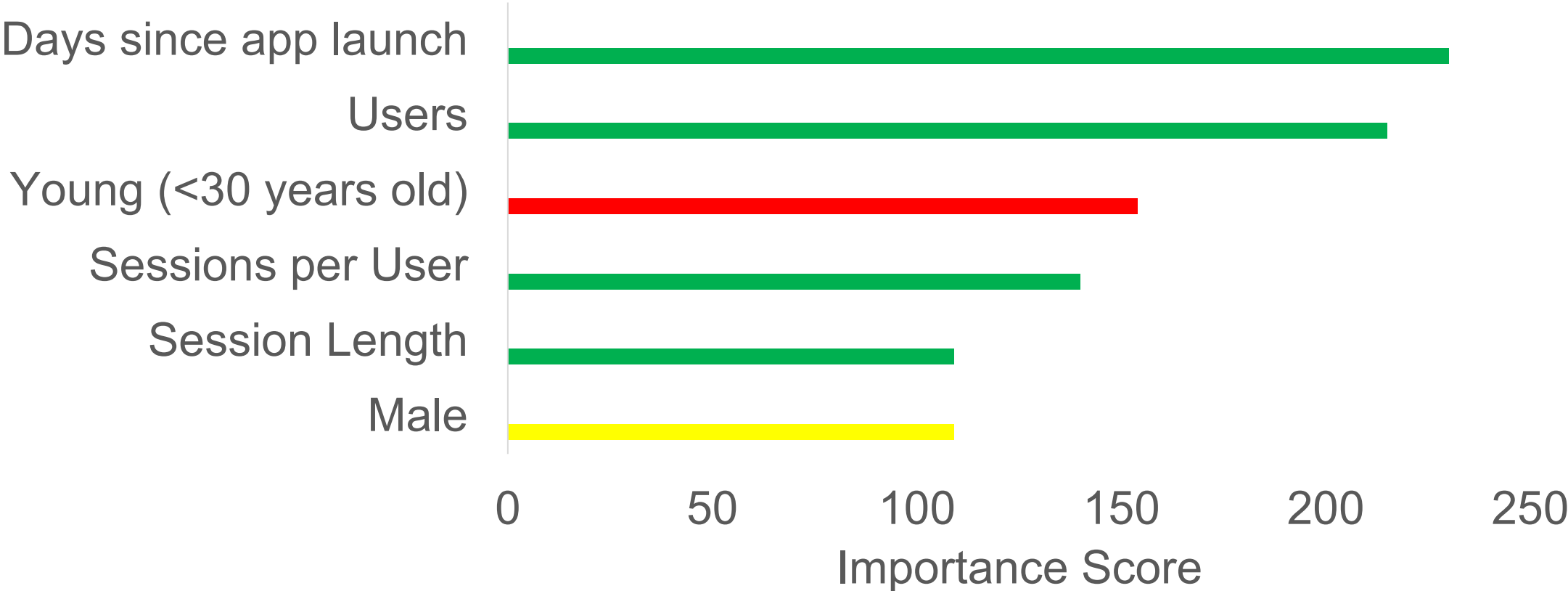
- 5-fold Cross Validation used (since test dataset is very small n=20)
- When using all features (averages over 5 folds):

Model	MSE (avg)	Accuracy	Precision	Recall
VAR	352.12	0.55	0.46	0.49
Polynomial Regression	306.44	0.53	0.45	0.45
Linear Regression	453.53	0.55	0.49	0.50
Prophet	337.71	0.51	0.43	0.41
Exponential (all coefficients)	235.31	0.49	0.44	0.44
Ensemble	210.41	0.55	0.47	0.45

# Results – Using only apps released since 2020



Most important features:



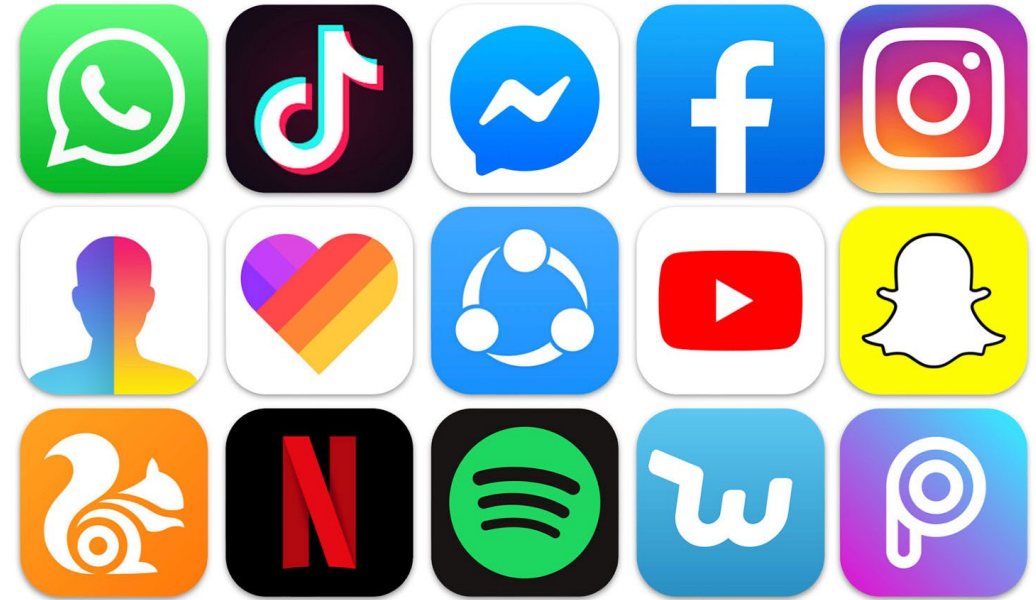
■ Positive Correlation   ■ Mixed Correlation   ■ Negative Correlation



# Conclusions



- Best model (overall) – ensemble
  - Lowest MSE and highest accuracy across 4 cuts
- Better predictions using first 30 days of data (up to 55%) vs random 30 days of data (up to 46%)
- Most important predictors of a future user base:
  - # of current users each day
  - % of older people who used the app



# Next Steps



- Use more complete dataset
- Use other time series models (e.g. neural network)
- Compare between app categories
- Try using only existing apps
- Use different prediction method

