



Data Driven Insights for Pro Bono Legal Services

DataFest 2023

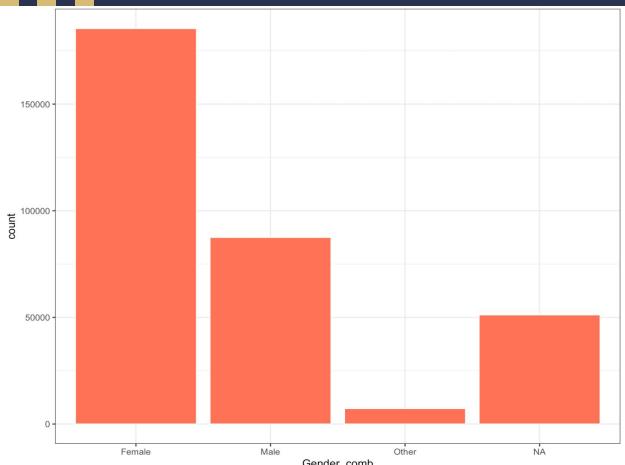
Team 3 / 04. 23. 2023

Our Research Question

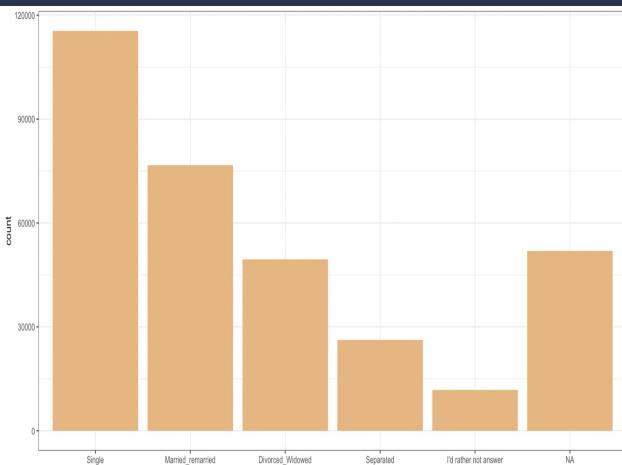


How can we improve the client experience of ABA Free Legal Answers?

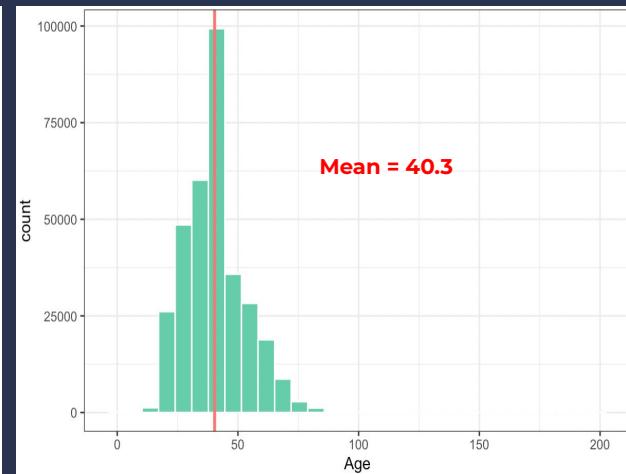
- Get **better insights of the clients** from their key demographics and geographical locations.
- Determine **the types of questions** asked by the clients.
- Analyze the **trends in the timing** of questions asked and in **turnaround time**.



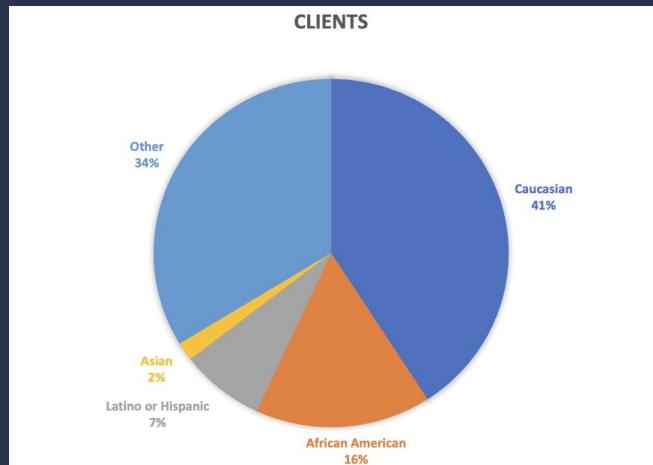
1. Number of Clients by Gender



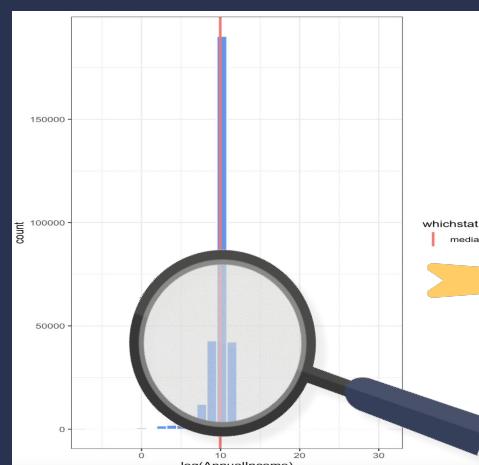
2. Number of Clients by Marital Status



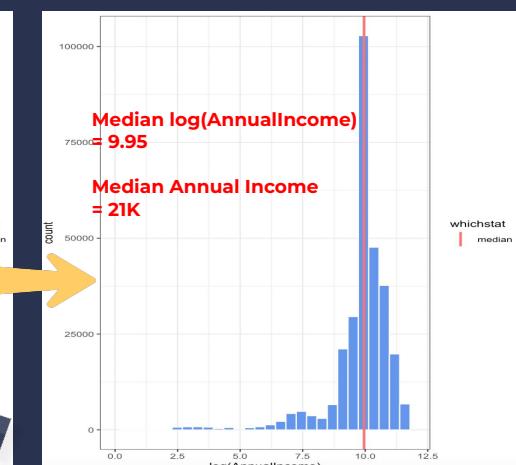
3. Distribution of Age



4. Ethnic Identity



5. Distribution of log(AnnualIncome)



Top 3 Potential Groups of Clients

Based on the Demographics



**Age 40-50
Annual \$ 20-40K
8.6% of the clients**

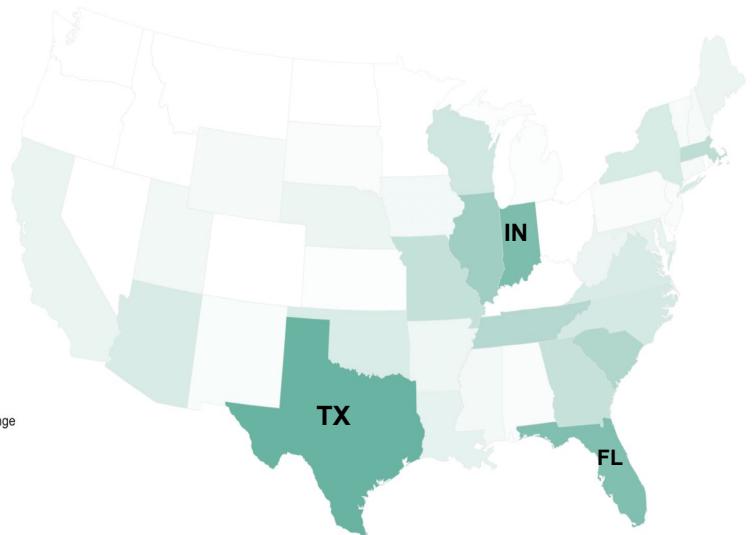


**Female
Caucasian
Single
Age below 30
Annual \$ below 20K**



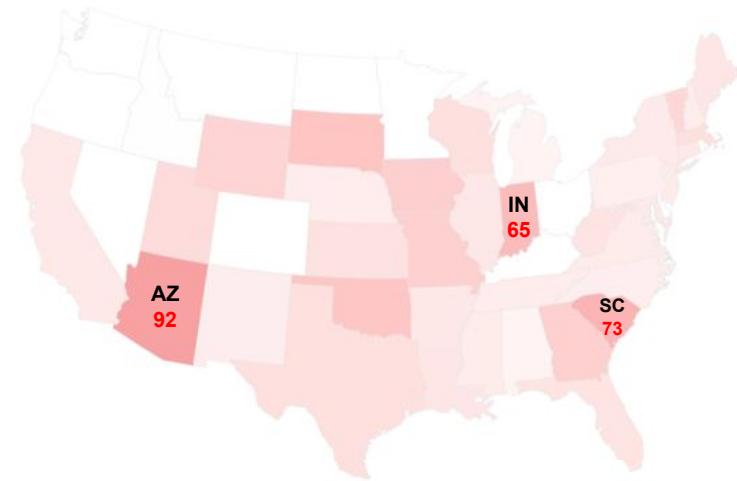
**Female
Caucasian
Divorced/Widowed
Age above 50
Annual \$ below 20K**

Client Population Distribution



High Demand
Texas, Indiana, Florida

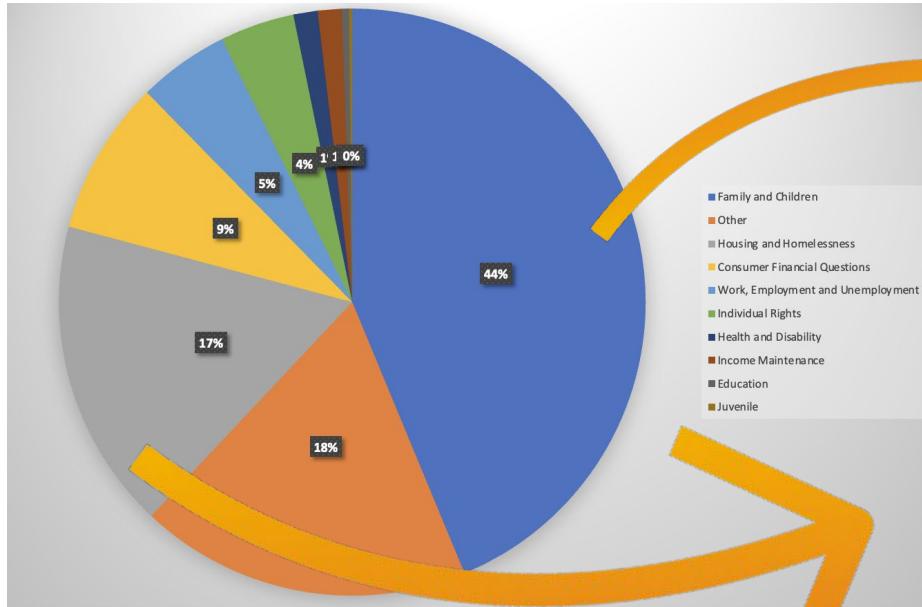
Average Number of Clients Per Attorney



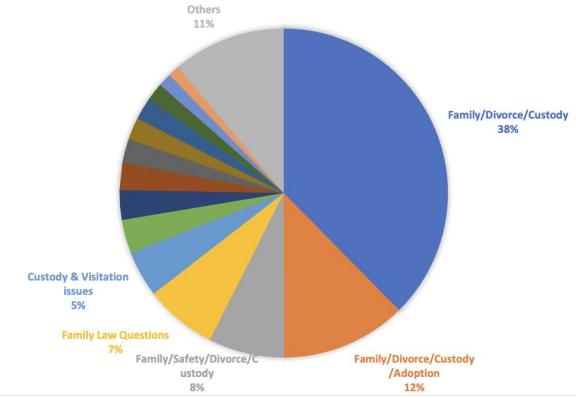
More Attorneys Needed
Arizona, South Carolina, Indiana

Category / Subcategory

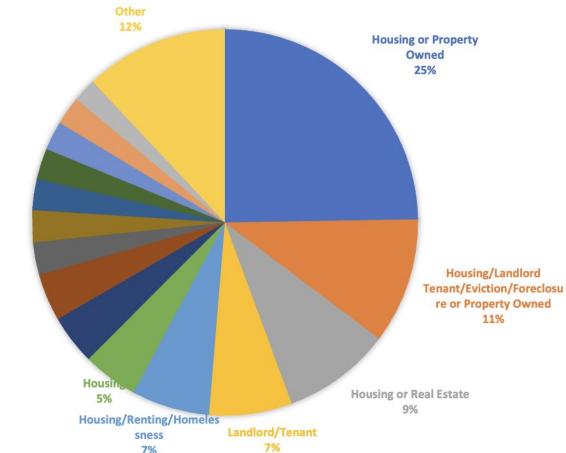
Category



Subcategory



Subcategory



Attorney / Clients Classifier

01

Classify 1000 posts

| | PostText | Label |
|------|--|-------|
| 1 | Do you still need help with this issue? It has been a little while since you posted. | 0 |
| ... | ... | ... |
| 1000 | Seems from all our calls to Troopers Animal Control etc that there is NOTHING AT ALL to be done. | 1 |



Label = 0: Attorney

Label = 1: Client

02

Feature Engineering & Modeling

03

Evaluate Random Forest Model

| | f1-score |
|----------|----------|
| accuracy | 0.91 |

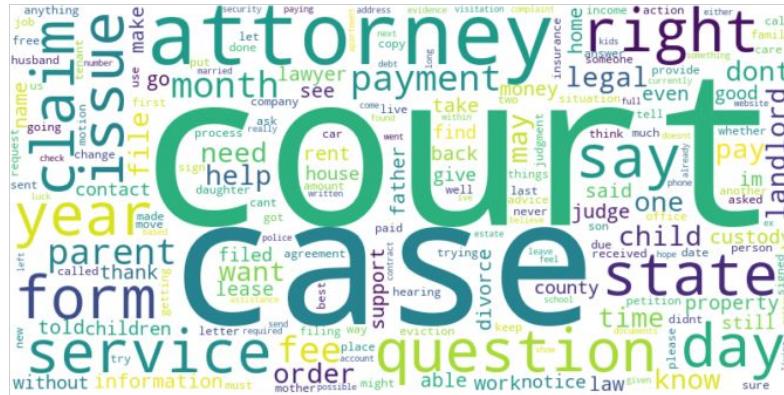
04

Classify on Full Data



Word Cloud

01. All Posts



02. Client Posts



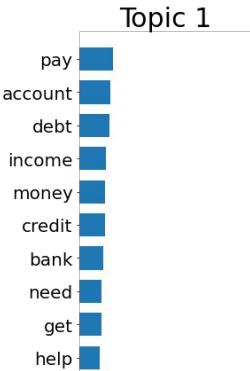
03. Attorney Posts



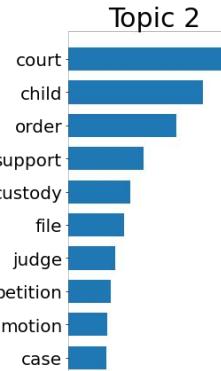
Topic Modeling

LDA(Latent Dirichlet Allocation) Model

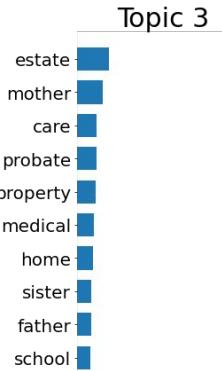
1. All Posts



Finance



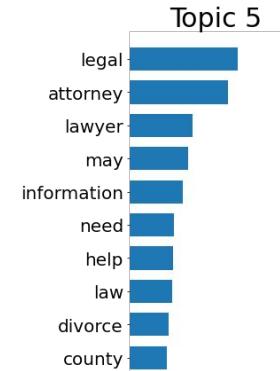
Child Custody



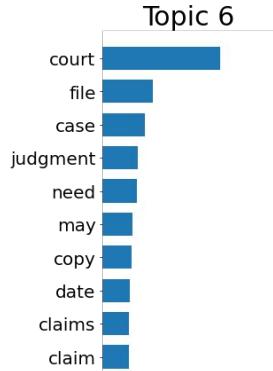
General



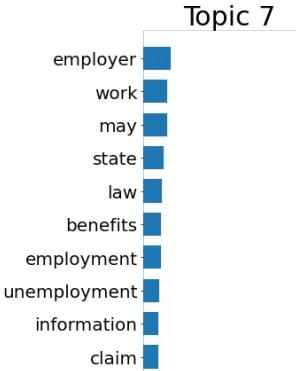
Landlord-Tenant



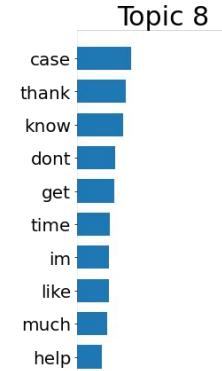
Legal Procedure



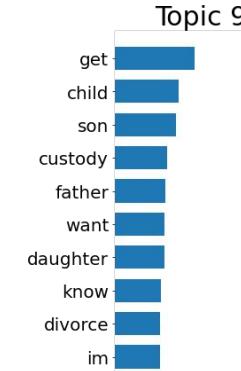
Court Procedure



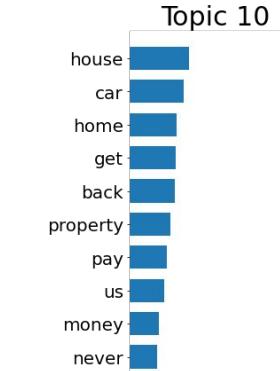
Employment



General



Family Law



Property transactions

Topic Modeling

LDA(Latent Dirichlet Allocation) Model

2. Client Posts

Topic 1

get
home
house
back
im
know
police
even
like
didnt

Topic 2

legal
attorney
lawyer
law
help
question
aid
may
free
services

Topic 3

property
account
agreement
contract
money
bank
lease
signed
name
house

Topic 4

child
custody
support
court
son
order
father
children
daughter
get

Topic 5

get
divorce
name
need
husband
know
want
help
house
dont

General

Legal Aid

Property transactions

Child Custody

Divorce

Topic 6

pay
insurance
car
company
vehicle
debt
income
amount
work
paid

Topic 7

court
case
file
filed
judge
attorney
hearing
need
motion
order

Topic 8

rent
lease
landlord
pay
us
month
back
get
time
apartment

Topic 9

thank
help
much
know
im
get
time
please
need
thanks

Topic 10

may
law
state
person
shall
medical
section
case
information
must

Finance

Court Procedure

Landlord-Tenant

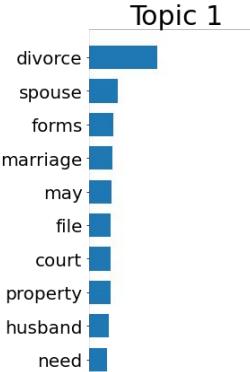
General

General

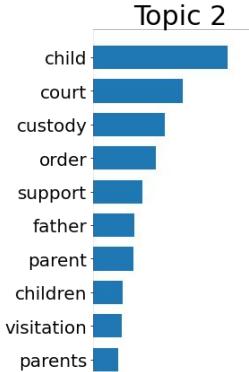
Topic Modeling

LDA(Latent Dirichlet Allocation) Model

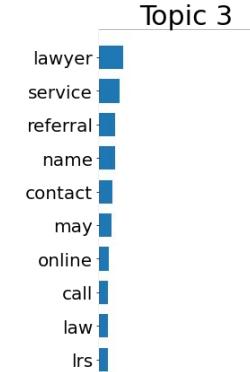
3. Attorney Posts



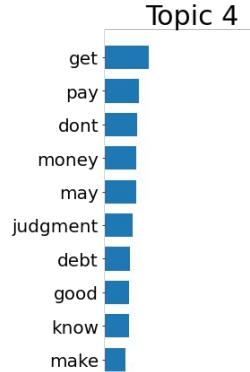
Divorce



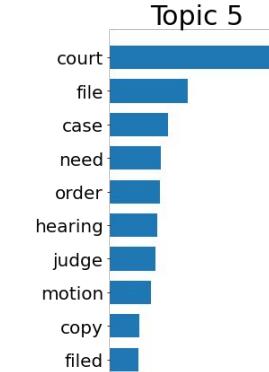
Child Custody



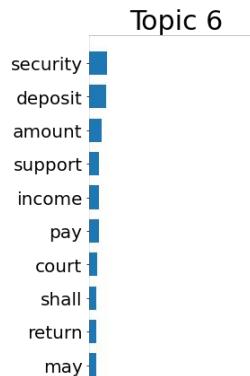
Legal Procedure



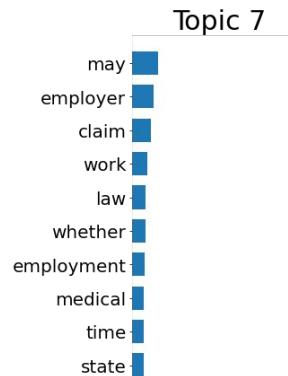
Finance



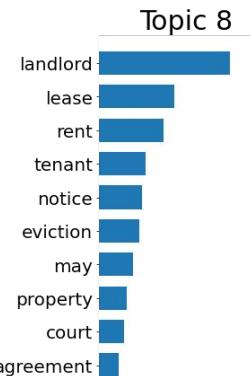
Court Procedure



Finance



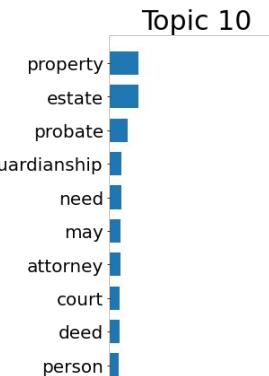
Employment



Landlord-Tenant

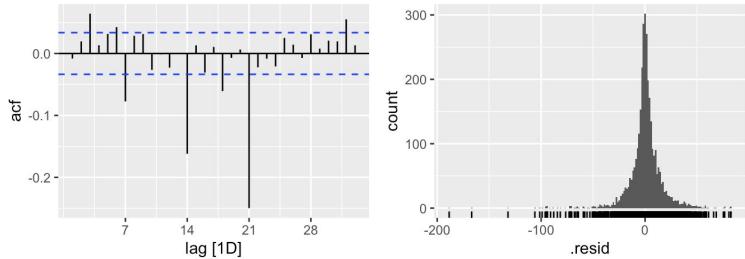
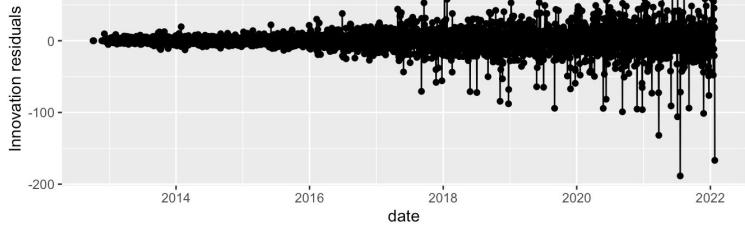
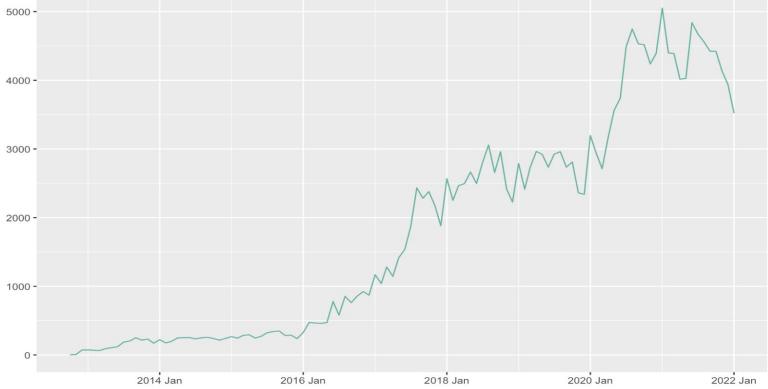


Legal Procedure



Property transactions

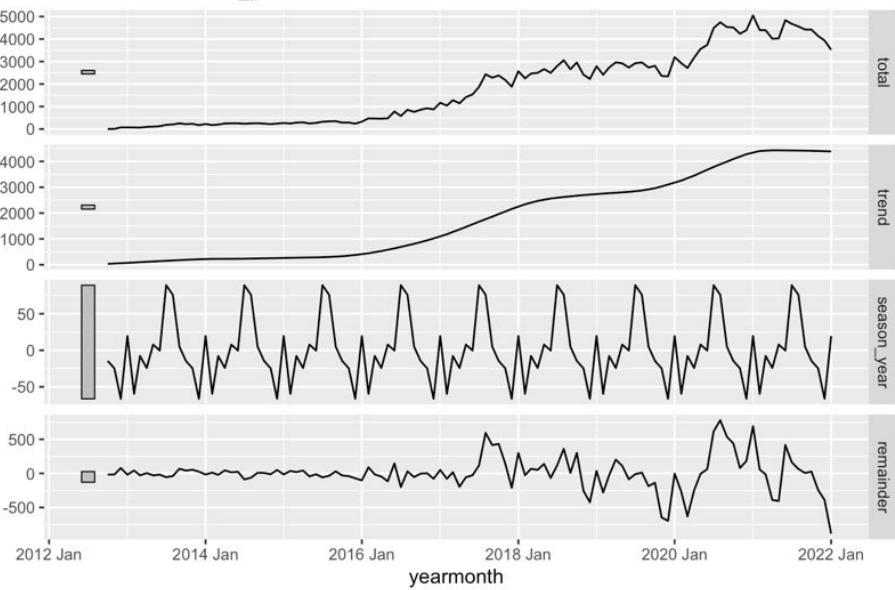
Number Questions through the time



Increasing trend with seasonal pattern

STL Decompose

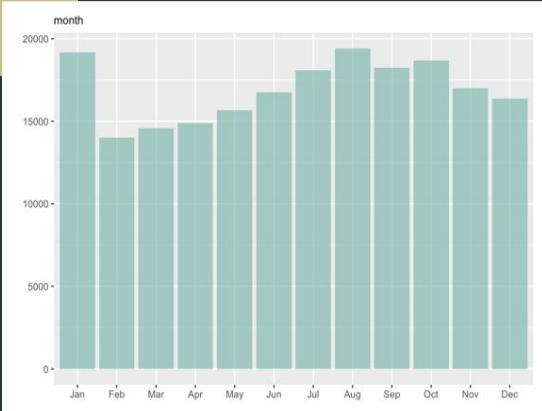
total = trend + season_year + remainder



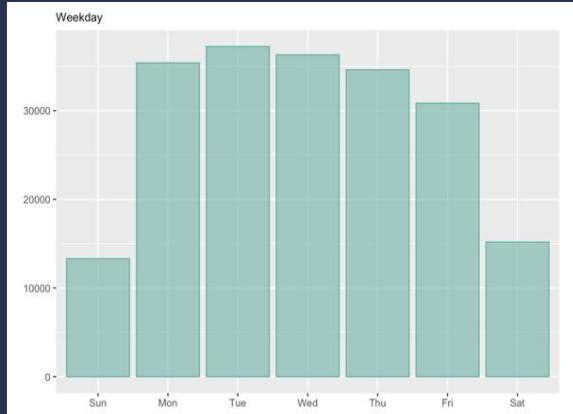
p = 1: one lagged value

ARIMA(1,0,0)(2,1,0)

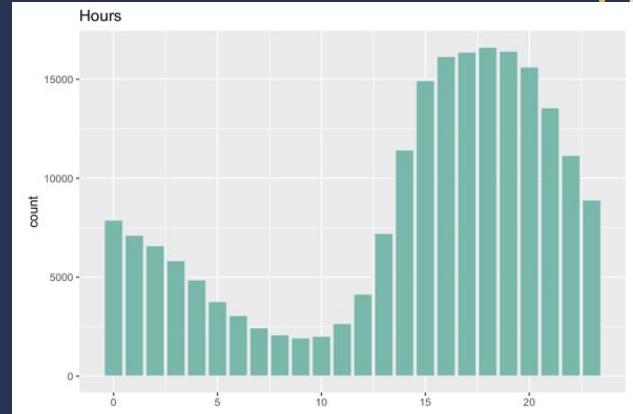
P = 2: two seasonal AR terms



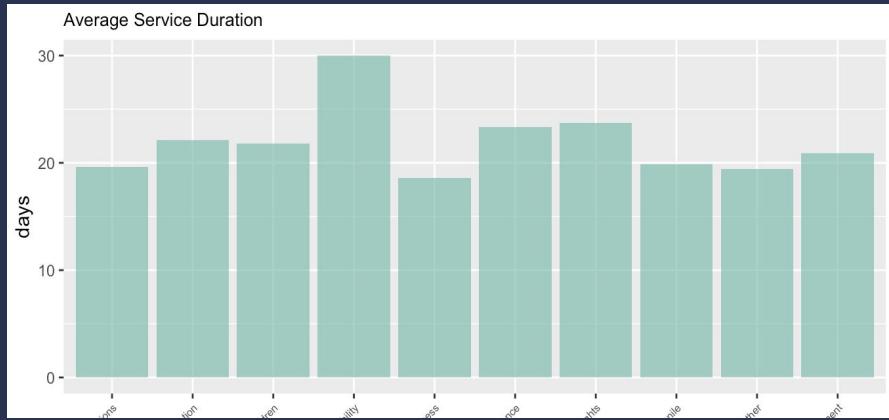
10-15% more volunteer should be added
on **Jan, Summer and Fall**



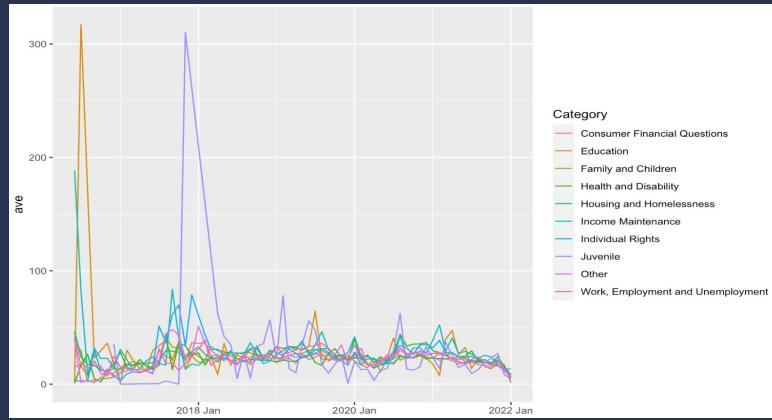
Business days mainly



Central time, peak volume from 1PM - 4PM



It takes average **20 days** per case,
Health and Disability take extra **10 days**

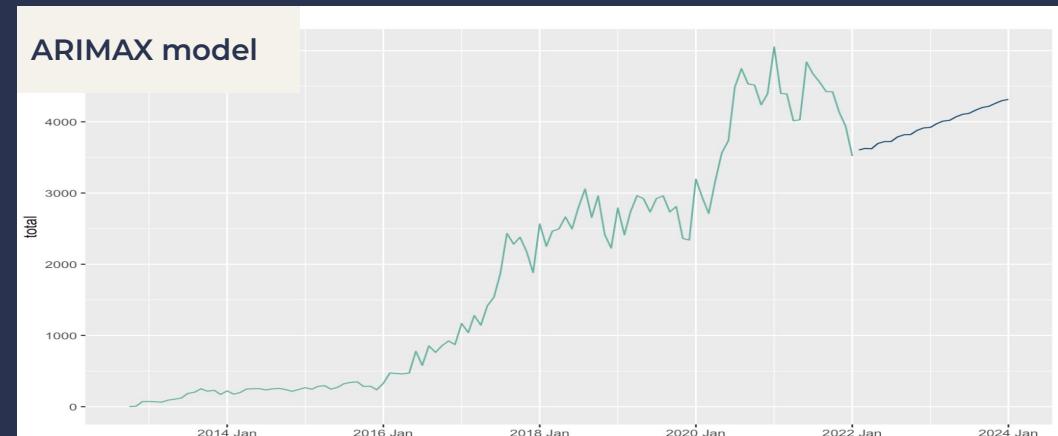
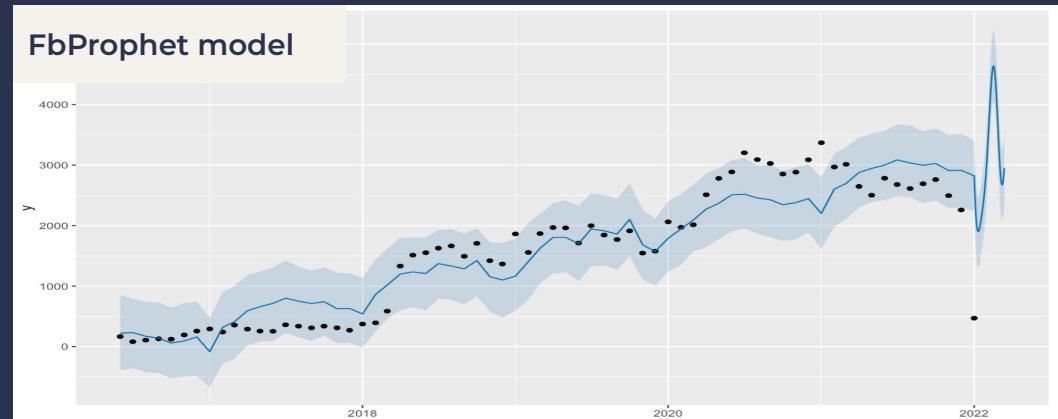


Before **2019** turnaround time for Juvenile and
Education case is 1 year in average.

Add new variable: Pandemic
Anova Test
 $P\text{-value} = 2.57e-11 < 0.05$

Prophet model suggest
there is a bounce back in beginning of 2022.

ARIMAX forecast in next two year volume will
increase from 3,500 to 4,300 questions monthly.



Suggestions

Large proportion of the clients are **middle aged with annual income \$20-40K and the types of questions** they're interested are

- **Family and Children**
- **Housing or Property Owned**

- Need to recruit more volunteers who have these **specializations**
- Need more volunteers **in Child Custody, Divorce, Property transactions topic**
- Need more volunteers **in January, Summer and Fall** based on our analysis in timing

Limitations

- PostQuestions is not divided into Question and Answer.
- Covid-19 intervened in time series data.
- There was no numerical response variable, so regression analysis was limited.

Tools



python™



R Studio®



Thank you !