

MAINE TRUST FOR LOCAL NEWS (METLN)

BEN COX, SARAH CRANE, ANDREW FUMAROLA

DS5110 – ESSENTIALS OF DATA SCIENCE, PROFESSOR BOCKMON

Geographic Analysis

Overview: Where are most news accounts?

- Total in Maine: 54608
- Total NOT in Maine: 5295

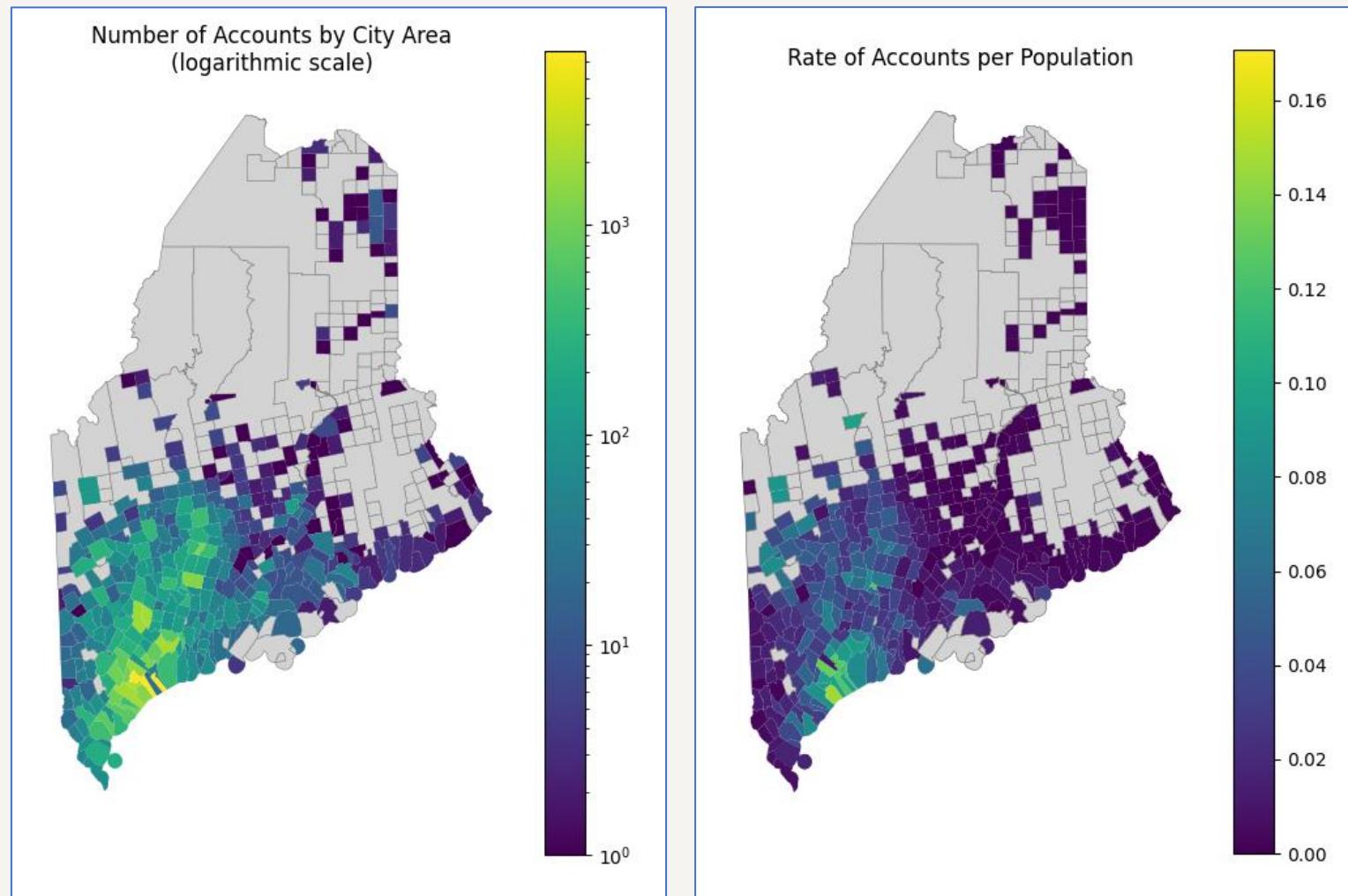
Only ~ 10% of subscriptions NOT in Maine.

Looking at Maine-specifically:

- High rates per population around city centers
 - (Portland, Lewiston, etc.)
- Lots of missing areas of data!
 - No accounts listed (mostly northern/rural)

Overall: High distinction between Southern & Southwestern Maine than the rest of Maine & borders areas

- Borders of this section are close to NH border: this area also has low subscription rates
- These areas may be best for outreach & growth

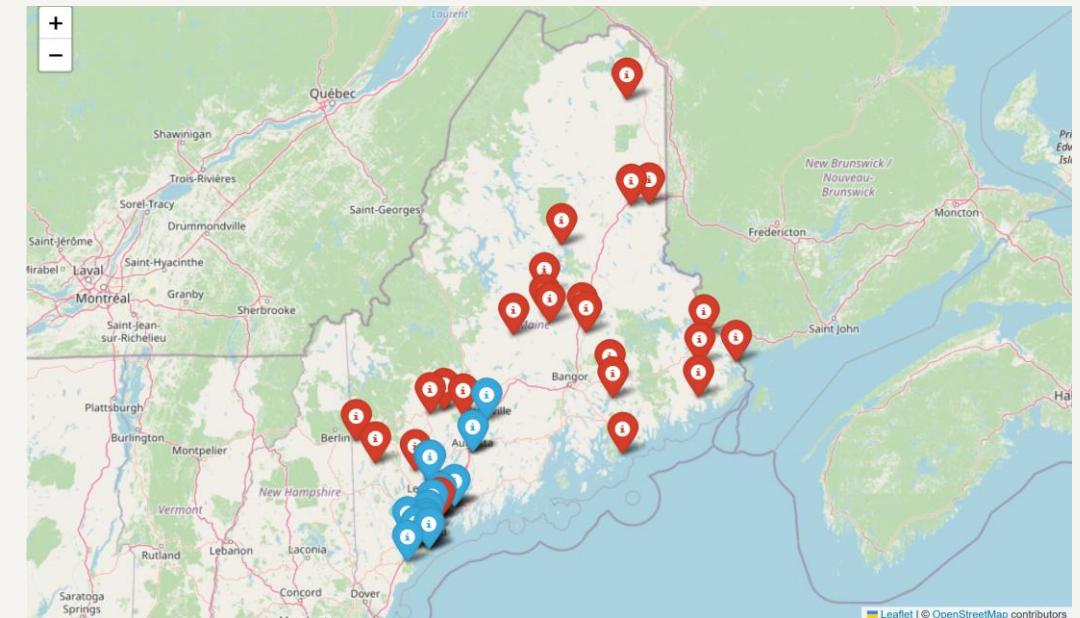


In this context, a "city" is a city, town, or other geographic area within ME .

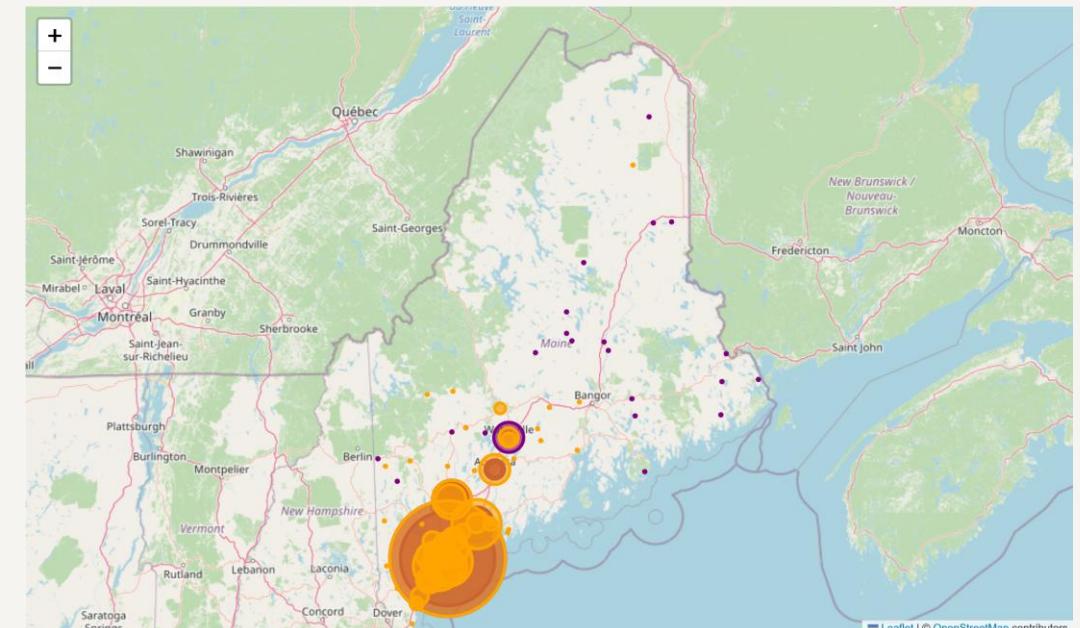
Rates = Number of Accounts in a city divided by Population of that city

Digital vs. Print

- Digital subscriptions made up 61.8% of the total subscriptions
- Maine accounted for 91% of subscriptions (all other national subscriptions were not examined further)
- Conclusions:
 - The major cities (Portland, Lewiston/Auburn, Augusta, Waterville, etc.) had the highest digital and print subscriptions
 - There is a noticeable shift in Waterville where digital starts to outweigh the print subscriptions
 - Rural communities appear more likely to rely on digital subscriptions



Highest and Lowest 25 Cities for Digital Subscriptions



Overlap of Digital vs. Print Subscriptions in Highest & Lowest 25 Cities

Temporal Trends

How do user numbers change over time?

How do new accounts and cancellations contribute to this DAU total?

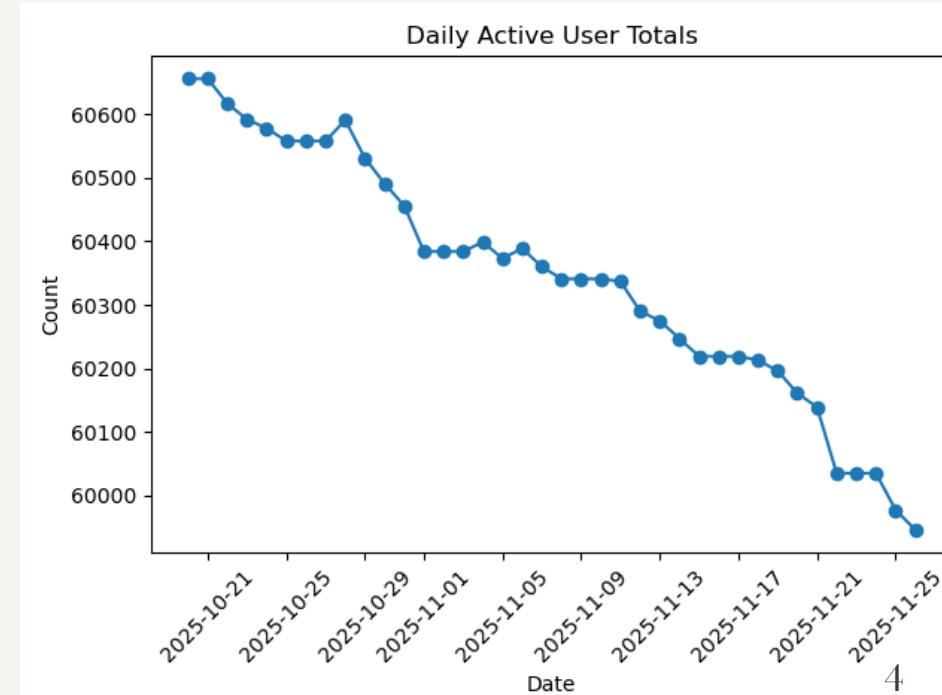
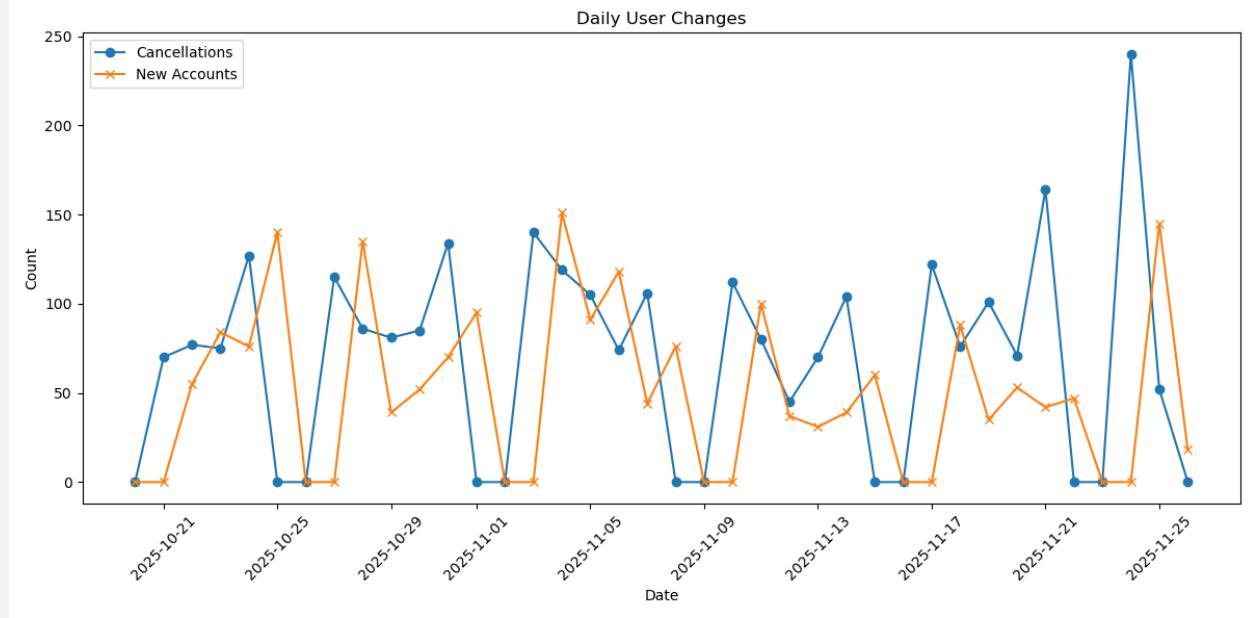
1. Compile daily logs and determine which entries are cancellations (last instance) or new accounts (first instance)
2. Aggregate number of cancellations and new accounts
3. Churn and Retention rates for given time period
4. View these counts by Zip, Publication, etc.

Viewing New Accounts by Publication

MTM_PT	1402
SMG_SJ	632
MTM_MS	391
MTM_KJ	45
AMG_TR	16

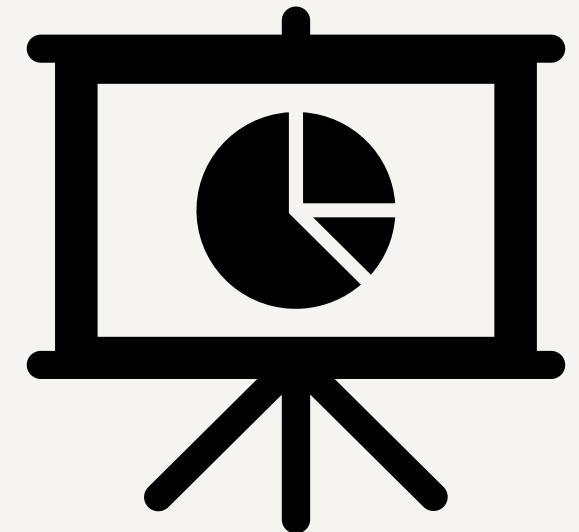
Viewing Cancellation Accounts by Zip

04240	191
04106	168
04103	155
04011	144
04210	132



Results

- Areas with the most potential for subscriber growth:
 - Mid-coast Region: Low subscription rates in these areas (per population).
 - Down East Maine: Few accounts or subscriptions present – area to expand.
- Rural communities favor digital subscriptions
 - Mid-coast region, Down East, NH border all favor digital subscriptions.
 - Print (or digital + print) still prevalent in Southern Maine, cities, urban communities.
- Over the data period 10/21 to 11/25
 - Churn Rate: 4.34%
 - Percentage of dropped subscriptions. The location with the highest churn was Freedom, ME with a churn rate of 18.75%
 - Retention Rate: 95.86%
 - Percentage of kept subscriptions (does not include new subscriptions)



Challenges



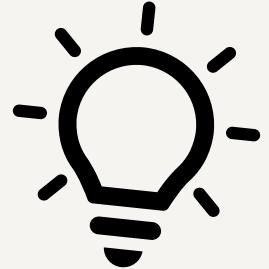
- Inconsistencies with naming and labeling within the data files
 - Additional pre-processing of data would be needed (ex. 'Portland' and 'PORTLAND' should be the same)
- Abbreviations in the naming of towns/cities
 - Some cities could not properly be geocoded due to shortening of their names (ex. 'Belgrade Lks')
- Data file (once combined as a stand-alone CSV) was harder to work with – inconsistent with how much of the data would be run while we were creating our visualizations
- Time – we had more ideas than we had time to tackle them...

Future Research

- Machine Learning Applications:
 - Using more historical data, could machine learning be used to predict which accounts are likely to cancel?
- Long-term time trends
 - What cycles contribute to churn and retention rates? (e.g. elections, annual cycle, major events)
- How do out-of-state subscribers contribute?
 - Is there a correlation with state migration?
 - Are there summer increases in subscribership, i.e. due to tourism or people with second or vacation homes?
- Investigating areas with higher education or schools – is there a pool of untapped subscribers?



Summary & Suggestions



- Our Goal:
 - Preliminary examination of the data regarding Geographic Analysis, Digital vs. Print Subscriptions, and Temporal Trends.
- What we found:
 - The Mid-coast, NH border, and Down East areas posed the most potential for subscriber growth
 - Digital & print subscriptions are popular within southern Maine, with digital being favored in rural communities.
 - Between 10/21 - 11/25 there was a Churn Rate of 4.34% and Retention Rate of 95.86%
- What would make future work with this data easier:
 - Being more selective with use of abbreviations
 - Better collection & organization of subscriber demographic information
 - Standardizing data with consistent valid entries for each variable (capitalization, spacing, etc.)
- Areas for further exploration:
 - ML Applications
 - Long-term trends
 - Out-of-state subscribers
 - School populations of untapped subscribers

Questions?



References

- Maine Trust for Local News, <https://www.metln.org/>
- Geopy, <https://geopy.readthedocs.io/en/stable/>
- Geopandas: <https://geopandas.org/en/stable/>
- Folium, <https://folium.readthedocs.io/en/latest/>
- Population estimates: <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>