...But Who's Complaining?

Demographic Influences on Corporate Responsiveness to Consumers

Jac Stewart, DS-SF-23, 7/7/2016

Contents

I. Motivating Question

Do **demographic characteristics** influence financial institutions’ **responsiveness** to individual consumer complaints?

II. Methodology

Research Question:

Which factors (consumer demographics, complainee-specific characteristics, or type of financial service offered) are most significant in predicting whether a given consumer user will receive a timely response to his/her complaint to the Consumer Financial Protection Bureau?

Key Terms

- **Complainant**: Consumer filing complaint
- **Complainee**: Subject of complaint
- **Verified Complaint**: Consumer complaint reviewed by the Consumer Financial Protection Bureau (CFPB) and found to reference a verifiable financial relationship between the complainant and accused financial institution (“complainee”)
- **Timely Response**: Whether or not the complainee responded to the complainant within fifteen days of being notified of a verified complaint

Data Sources & Characteristics

- **Consumer Complaints**:
  - Source: Consumerfinance.gov
  - Date Range: July 2013 - June 2016
  - Overall Size: 590,081 user complaints
  - Geographic Representation:
    - Zip Codes: 27,313 (out of ~43,000 across United States)
    - States: 63*

- **Income Data:**
  - Source: IRS.gov
  - Date Range: 2013
  - Geographic Representation:
    - Zip Codes: 20,706
    - States: 51*

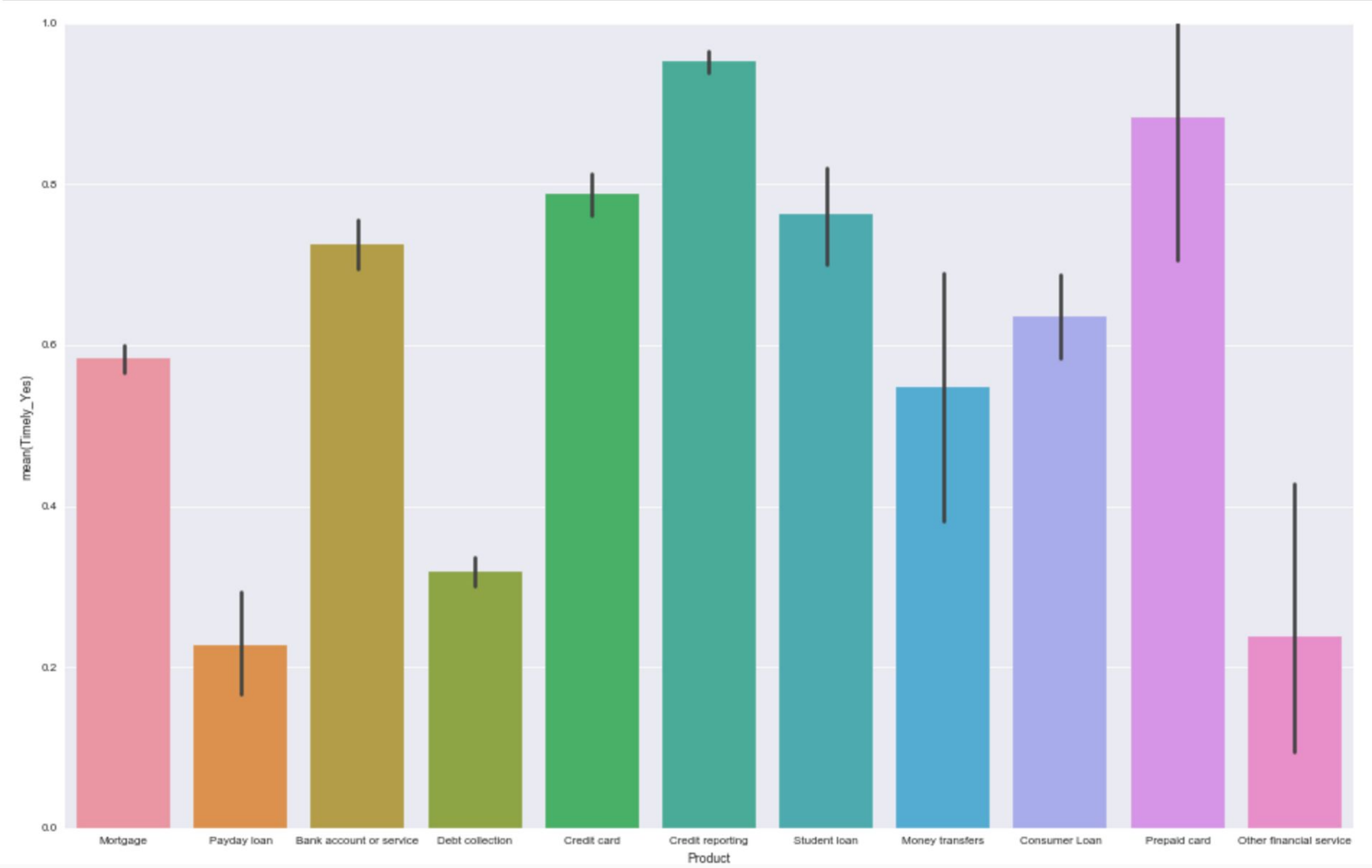
* The additional 12 states not used in my analysis include values I was unable to disambiguate (e.g., FM), which correspond 1:1 to those lacking IRS income data. Some, such as "PR," and "GU," may refer to U.S. territories, such as Puerto Rico and Guam.
** I count Washington, D.C. as a 51st state

Methodology:

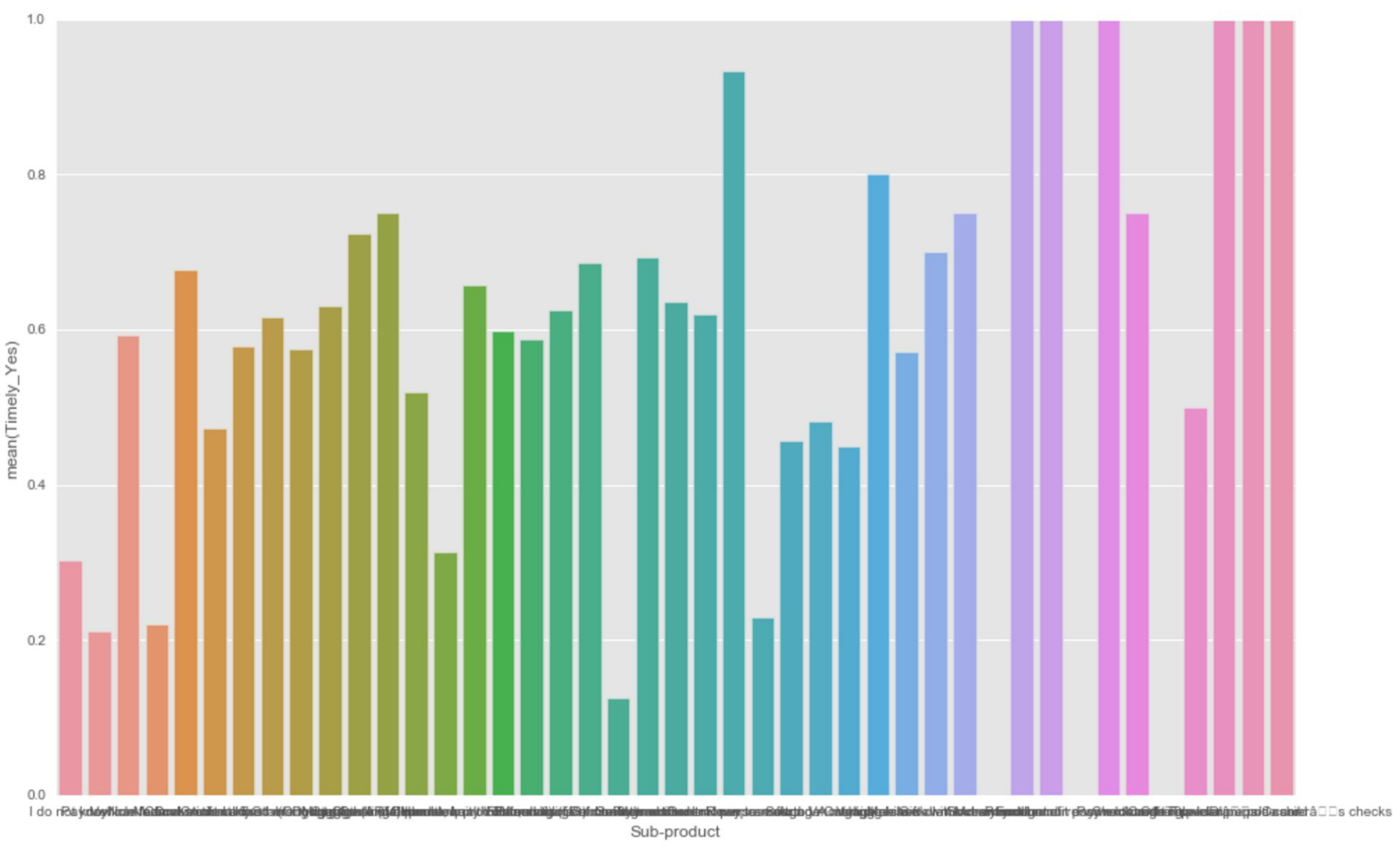
- **Logistic Regression**:
  - Binary, categorical dependent variable (DV)
  - Categorical and continuous independent variables (IVs)
    - Categorical variables converted to dummies

- **Variables Considered**:
  - Financial Service
    - Product (e.g., “bank account or service”)
    - Sub-Product (e.g., “account status”)
  - Issue Type
    - Issue (e.g., “cont'd attempts collect debt not owed”)
    - Sub-Issue (e.g., “frequent or repeated calls”)
  - Social Status
    - Senior, Veteran, Senior & Veteran
  - Adj. Gross Income
    - State-Level
    - ZIP Code-Level
  - Submission Method

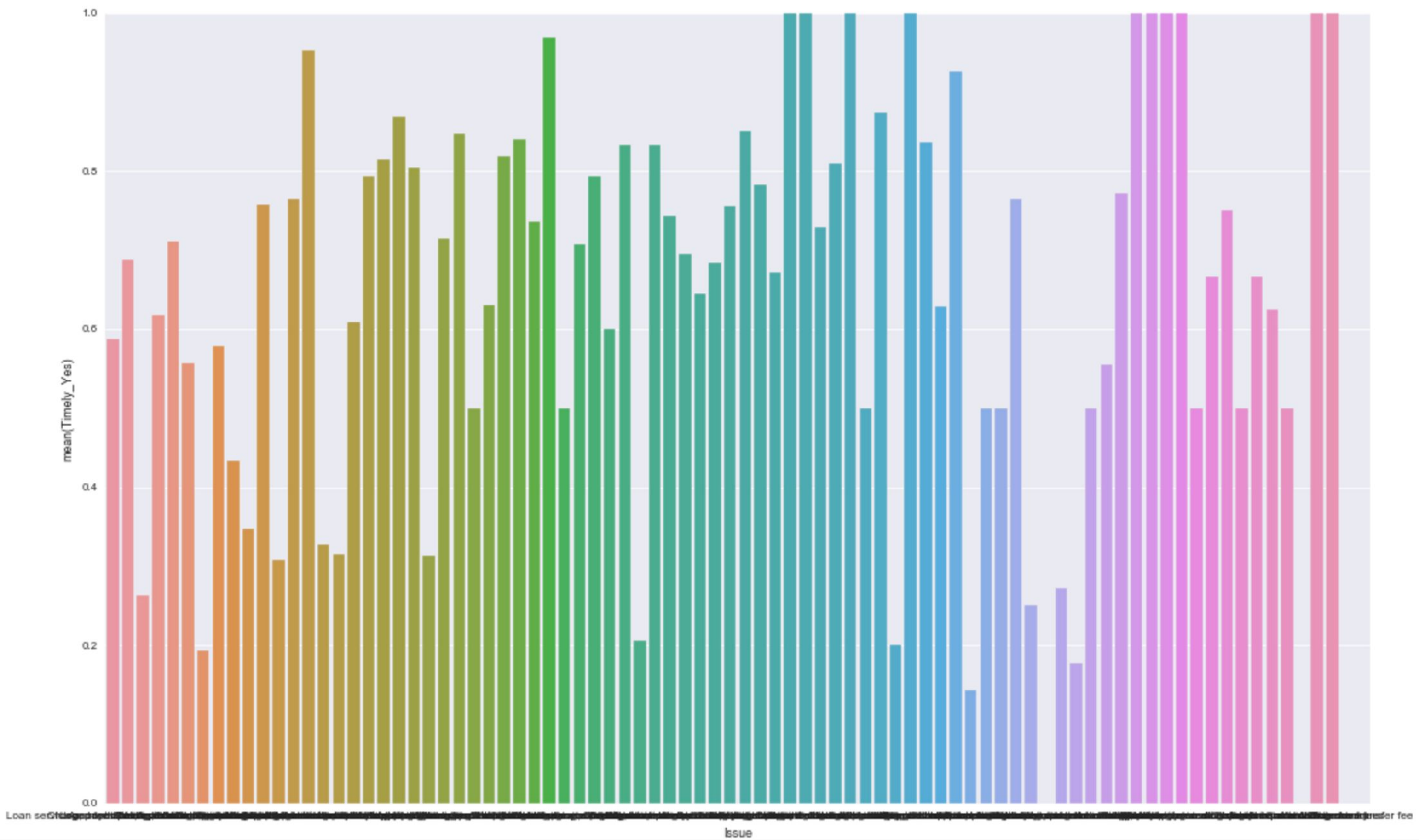
Product vs. Timeliness (Mean)



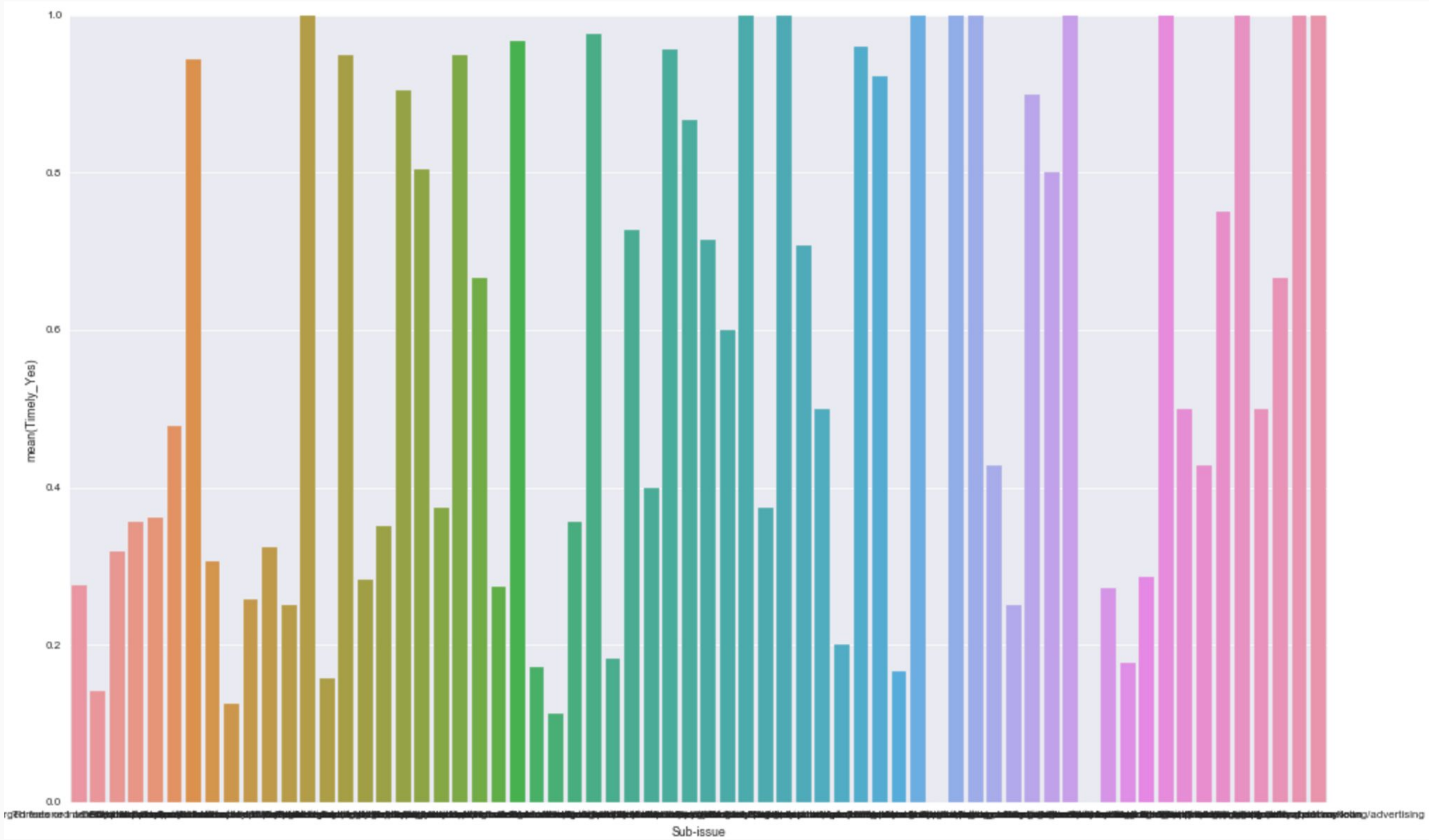
Sub-Product vs. Timeliness (Mean)



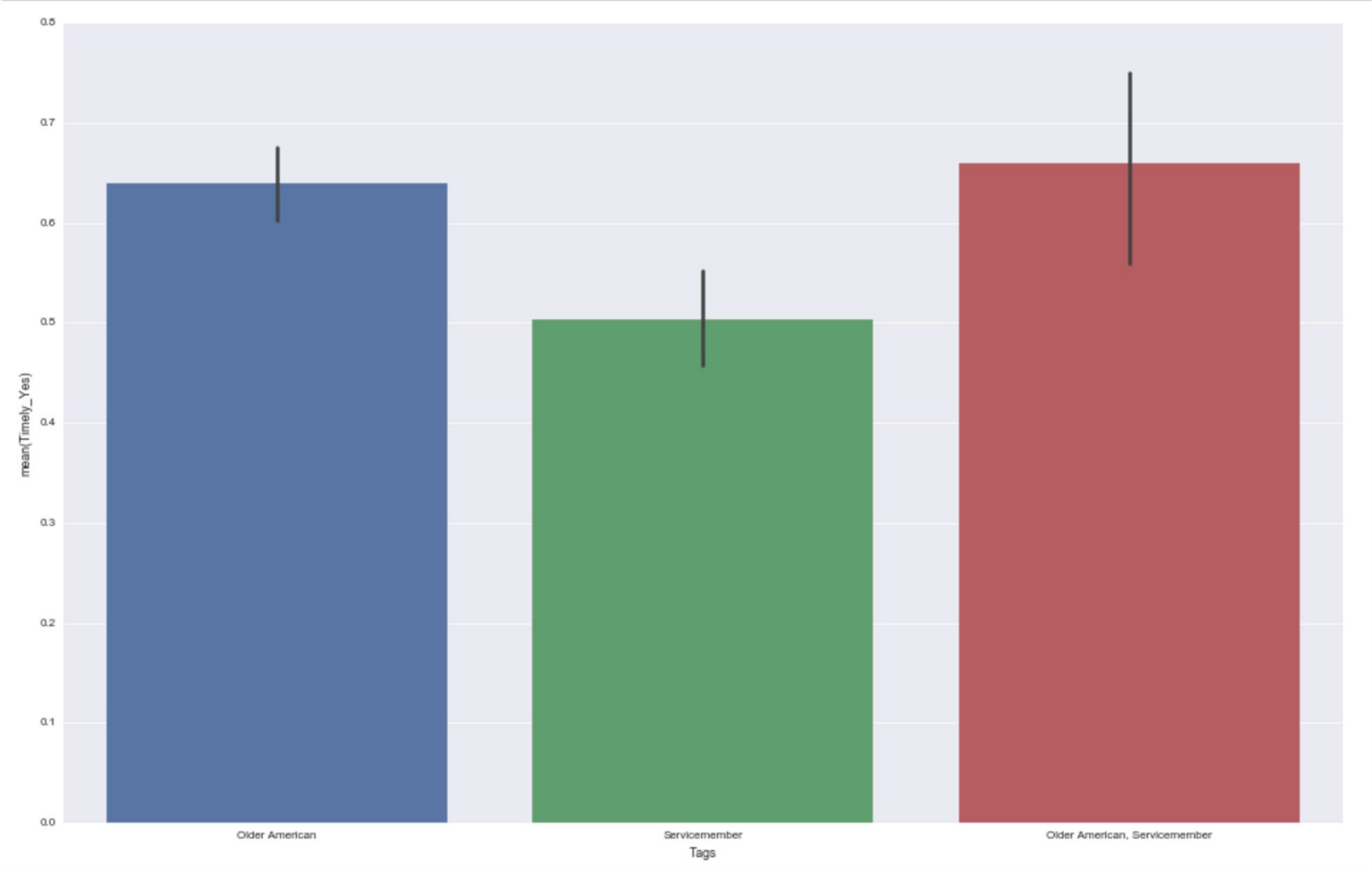
Issue vs. Timeliness (Mean)



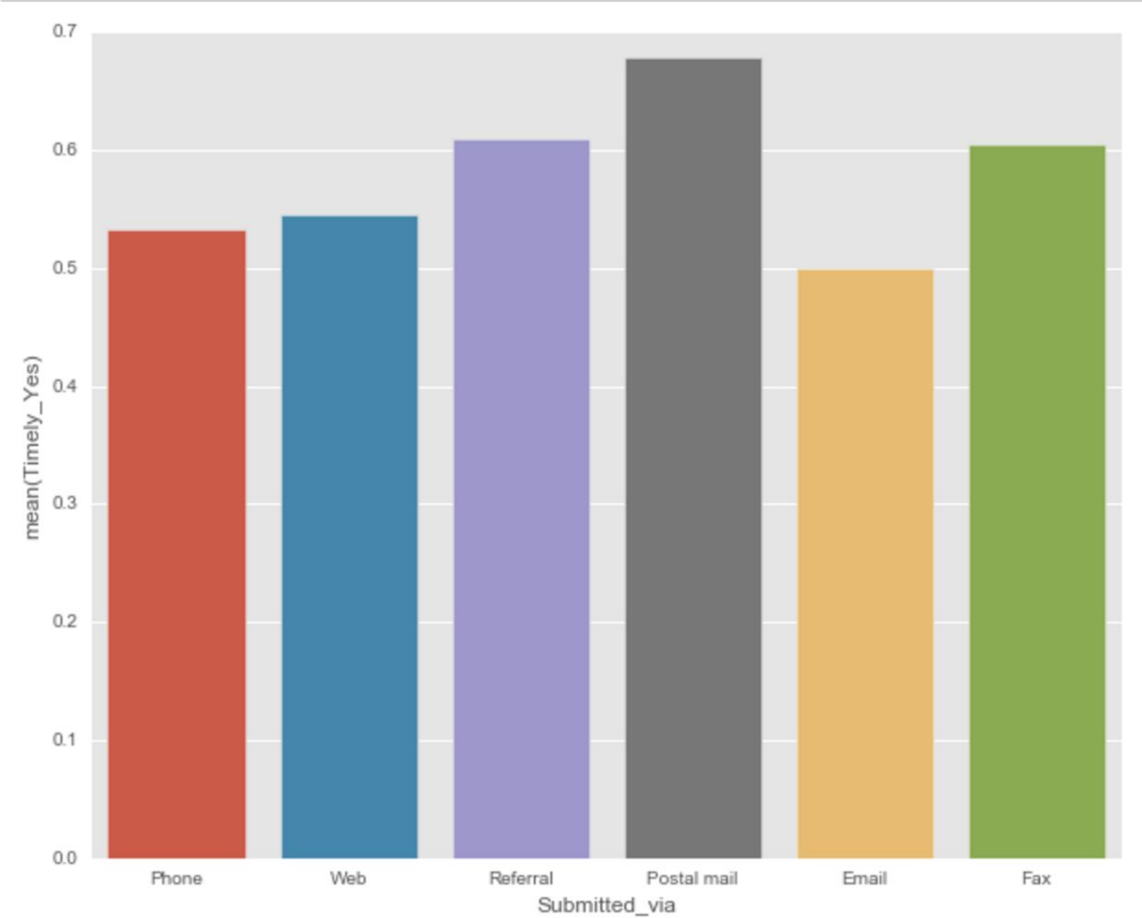
Sub-Issue vs. Timeliness (Mean)



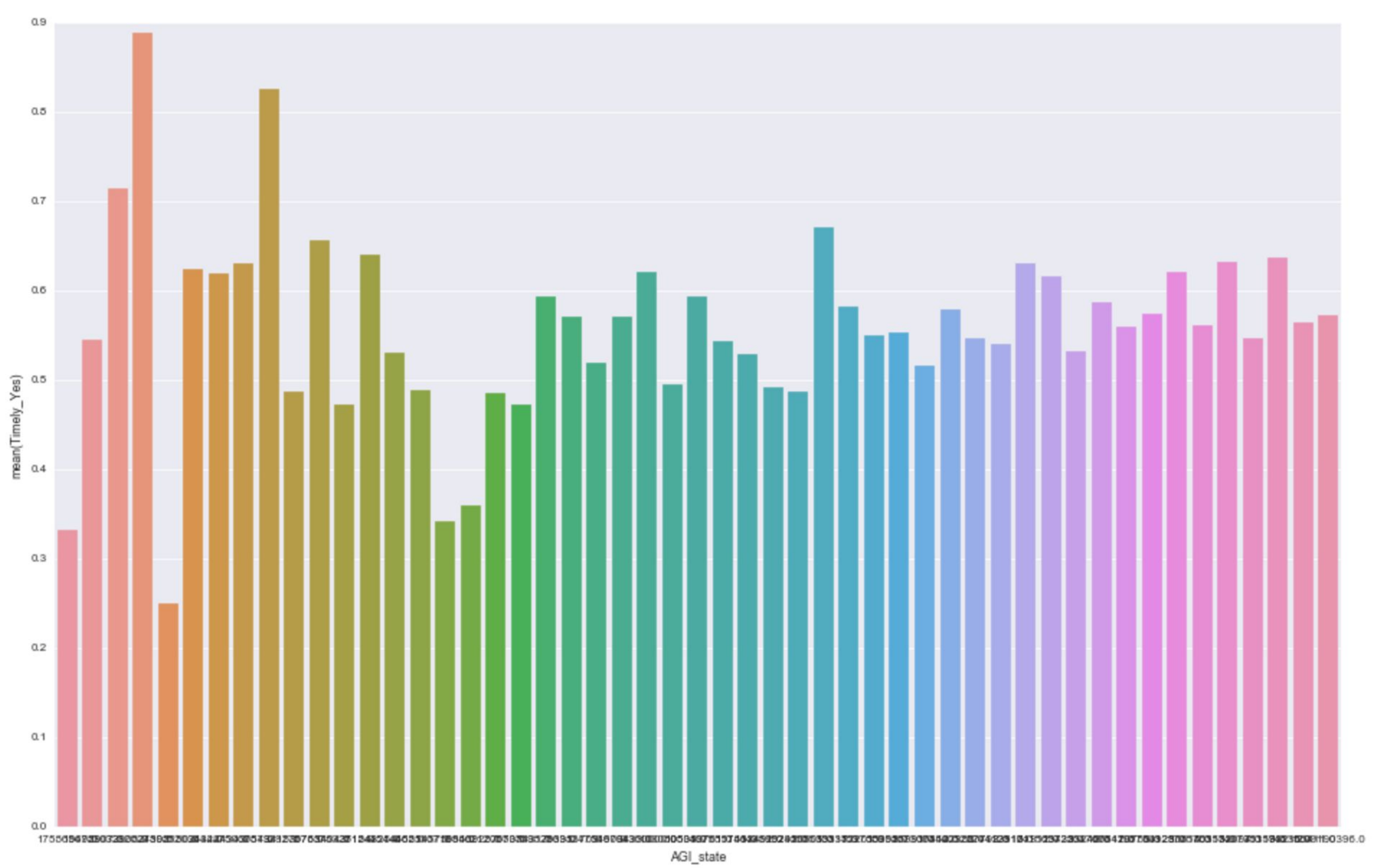
Social Status vs. Timeliness (Mean)



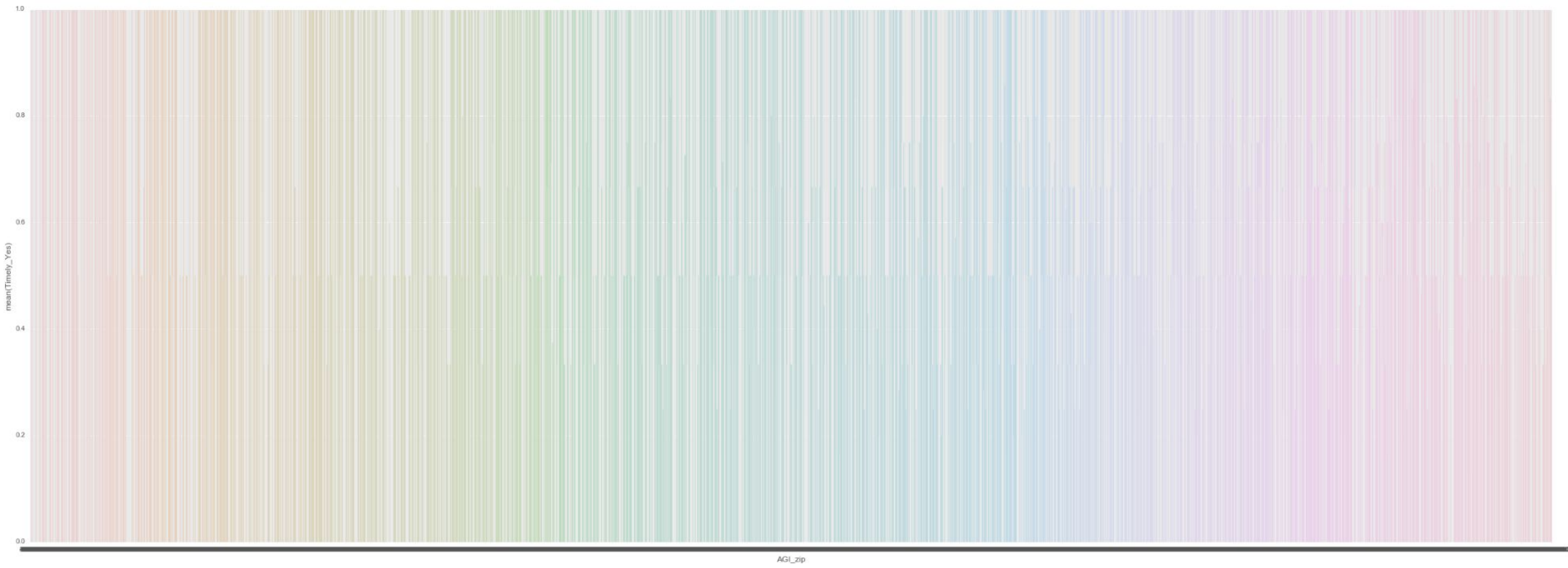
Submission Method vs. Timeliness (Mean)



State-Level Income vs. Timeliness (Mean)



ZIP-Level Income vs. Timeliness (Mean)



Hypothesis

Top predictors of response timeliness will be:

1. **Product**
    a. Avg. case complexity (e.g., “bank account” vs. “credit reporting”)
    b. Number of parties involved
2. **Income Level of Complainant’s ZIP Code**
    a. Mix of financial services offered/used in low- vs. middle- vs. high-income areas
    b. Differences in incentives for developing customer service infrastructure based on business model
3. **Issue**
    a. Avg. case complexity (e.g., “unsolicited credit card” vs. “identity theft”)
    b. Number of parties involved
    c. Signal may be weakened by:
        i. High number of categories
        ii. Subjectivity of classification

III. Results & Analysis

Final Model & Results

- **Model:**
  - Products:
    - All but one dummy ('Other financial service')
  - Social Status
    - Single dummy for non-veteran seniors

- **Predictive Accuracy:**
  - Training Data:
    - R^2: 0.14
    - P-value Per Dummy: < 0.01
    - Sklearn LR Value: 0.684
  - Test Data:
    - R^2: 0.13
    - P-Values < 0.025:
      - Prepaid Card (0.04)
      - Senior (0.026)

Hypothesis Evaluation

Top predictors of response timeliness:

1. **Product**
   - Most predictive categories (z-scores):
     i. Debt Collection (-19)
     ii. Credit Reporting (19)
     iii. Bank Account or Service (13)
     iv. Credit Card (13)
2. **Income Level of Complainant's ZIP Code**
   - Overly segmented data
3. **Issue**
   - Overly segmented data

IV. Executive Summary

Executive Summary

- No evidence of direct relationship between complainant demographics and corporate responsiveness

- Evidence of **indirect relationship** through **differences by industry**

- Payday loan and debt collection industries drive untimely response volume, raising concerns about indirect disparities in service level for lower-income Americans

Assumptions & Future Research Opportunities

- **States Examined**: Limited to zip codes from IRS data
- **Role of Complainee Characteristics**:
  - Too many unique companies (3334) to model
  - Categorization based on additional data may add signal:
    - Untimely response rate
    - Governance structures (e.g., public vs. private)
    - National vs. local
- **Role of Time-Related Characteristics**:
  - Did not examine factors such as :
    - Secular changes in timeliness over time
    - Seasonality / cyclicality of complaint volume

Assumptions & Future Research Opportunities

- **Sub-Product and Sub-Issue Classification**:
  - Too many categories (40 and 66, respectively)
  - Grouping categories may isolate signal
    - Ex: “billing dispute” and “don’t agree with fees charged”
- **Income Classification**:
  - Too few categories at State level (51) and ZIP level (~20,000)
  - Grouping ZIPs may isolate signal
    - Ex: Assigning values based on income quartile
- **Additional Modeling:**
  - Given ultimate model, could test KNN
  - Random Forest may reveal unanticipated relationships

Appendix

See iPython Notebook (and code to be posted on Github) for full documentation!

Thank you!
Questions?