



CMSC 12300

Computer Science with Applications III

By

Thomas Wilson, Matthew Auerbach and Tyson
Miller

Exposing Crime on Venmo

Dataset

- There is no public dataset of all venmo transactions. Our dataset would need to be scraped from Venmo's public API
 - <https://venmo.com/api/v5/public?limit=1>
- Although we are not supposed to scrape to obtain our dataset, it does not seem that complex/difficult
- It is hard to get a sense of how large the dataset will be without scraping it for ourselves.
 - We drew inspiration from <https://publicbydefault.fyi/>—a project that analyzed venmo transactions in 2017 (of which there were ~ 207,000,000)
 - I imagine we will have to limit our analysis to a certain time period to limit the size of the dataset
- Entry Example

```
- {
  payment_id: 1888069632,
  permalink: "/story/5cbe403e7addfb4bc77543bf",
  via: "",
  action_links: { },
  story_id: "5cbe403e7addfb4bc77543bf",
  comments: [ ],
  updated_time: "2019-04-22T22:29:18Z",
  audience: "public",
  - actor: {
    username: "trentwhite15",
    picture: "https://venmopics.appspot.com/u/v1/m/eeel4003-0eaf-4e14-a097-7cc85e41b11e",
    is_business: false,
    name: "Trent White",
    firstname: "Trent",
    lastname: "White",
    cancelled: false,
    date_created: "2017-09-04T05:19:14",
    external_id: "2296250191839232368",
    id: "25636685"
  },
  - transactions: [
    - {
      - target: {
        username: "justinnolan34",
        picture: "https://venmopics.appspot.com/u/v3/m/2ef142b7-ffe2-443f-9480-c2d3ca9e76f8",
        is_business: false,
        name: "Justin Nolan",
        firstname: "Justin",
        lastname: "Nolan",
        cancelled: false,
        date_created: "2017-07-28T01:36:12",
        external_id: "2268596457701376129",
        id: "23935552"
      }
    }
  ],
  created_time: "2019-04-22T22:29:18Z",
  mentions: [ ],
  message: "👉"
```

Goal

The goal of our project is to leverage the username, message and transaction amount fields to try and uncover recurring illegal payments. In order to accomplish this we would need to employ

several heuristics. One possibility would be to look through messages qualitatively, seeking out keywords such as “drugs”, “pot” or “~leaf emoji~”. Another possible heuristic is to look for individuals with recurring payment amounts or for individuals with an above average weekly or monthly transaction number. Once we find obvious criminals, we can search through their transactions to develop more heuristics and search the dataset again with the additional parameters.