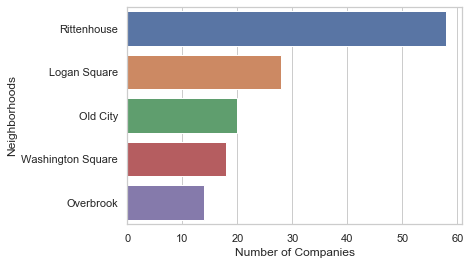
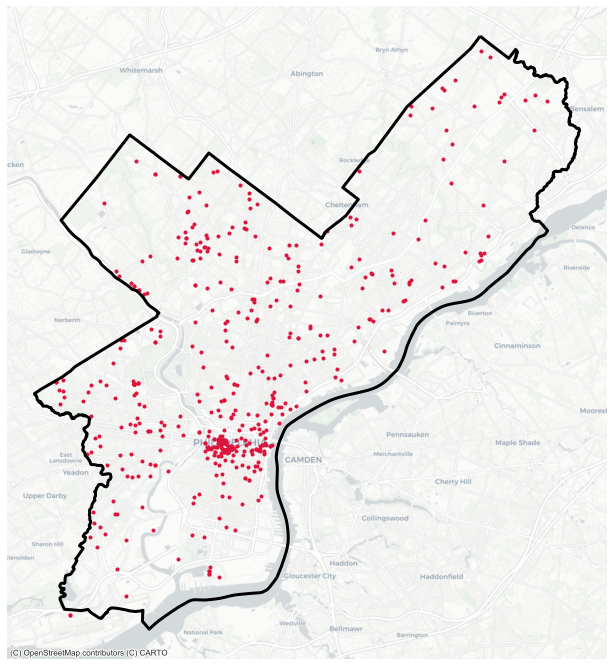
**ANALYSIS ON MWDSBE**

*Office of The Controller*

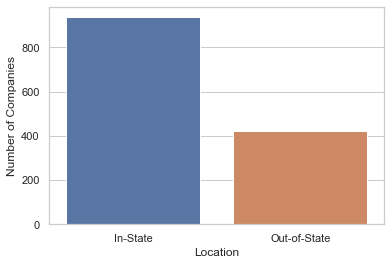
* **Data**
  + Registry: all registered MWDSBE
  + License: Commercial Activity Licenses. Information includes license information such as license number, activeness, etc.
  + State License: list of sales tax licenses, exemption and wholesaler certificates as maintained by the Department of Revenue (DOR) that are active
  + Payments: includes checks and ACH (direct deposit) payments made by the City during 2017 fiscal year.
  + Professional Services: Professional Services Contracts by fiscal quarter from 2013 to 2019
* **Merging Algorithms**:
  + Fuzzy Wuzzy with score cut-off 95
  + TF-IDF with score cut-off 85
  + Execution Time
    - Fuzzy Wuzzy: about 20 minutes
    - TF-IDF: about 3 minutes
* **Registry + License**
  + Matched: 1566 out of 3119
  + Active companies: 1553 out of 1566 (99%)
  + Top 5 Neighborhoods



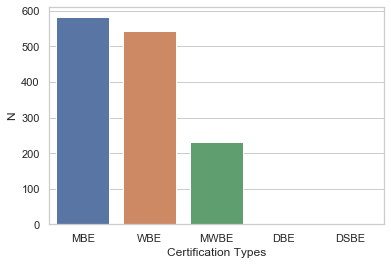
* + Location of MWDSBE in Philadelphia



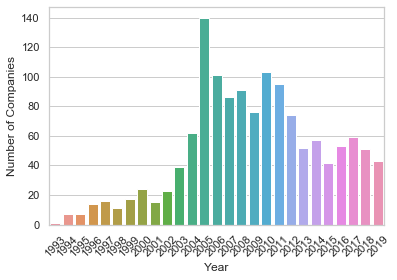
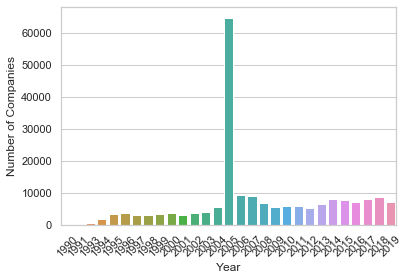
* + In-State vs Out-of-State / In-City vs Out-of-City



* + Certification Types

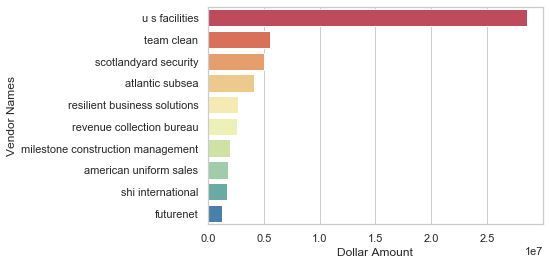


* + Year Trends for Registry vs matched

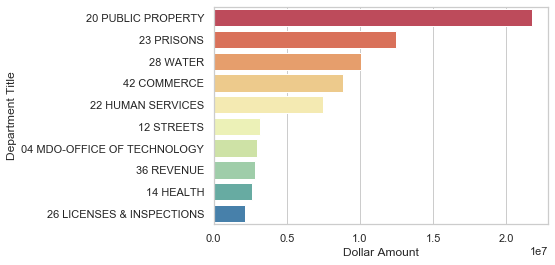


*\* NOTE: only used TD-IDF for fast testing*

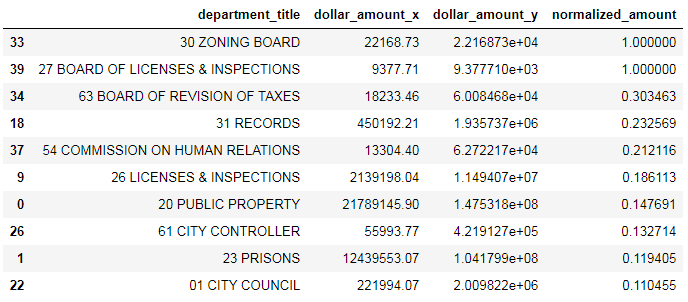
* **Registry (only in PA) + State License Data**
  + Matched: 564 out of 1769
    - Newly found companies: 201
* **Registry + Payments**
  + Matched: 369 out of 3119
    - Newly found companies: 61
  + Top 10 Vendors by payments



* + Top 10 Departments by payments



* + Normalized Transaction Amount of MWDSBE by departments



* **Registry + Professional Services Payments**
  + Matched: 205 out of 3119
    - Newly found companies: 71
* **Current Total Matches** = 1566 + 201 = 1767 out of 3119 (56.6%)
  + Only applied to “Registry + License”. Applying Fuzzy and TF-IDF algorithms to all dataset will increase the number (estimate 60%).