Data Description

The overall sample consist in 900,000 licenses, from the years 2001 until 2019. We convert this sample into unique business histories using the account number which identify it. Doing this manage to obtain around 200,000 unique histories in the data set. Afterwards, also we select the companies who start in the period of analysis. For doing this we select the licenses issued in the year t. Also, we drop the observations that have less than 10 days. On overall, the whole data set have 150,470 unique business. Additionally, we use other sources as the Chicago Crime Data and the transit data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data** | **Description** | **Time available** | |
| Business Licenses | Data base with characteristics of the licenses | | 1998-2019 |
| Chicago crime data | Data base with the reported crime data | | 2001-2019 |
| Transit data | Transit data of Chicago | | 2006-2019 |
| Chicago Bounderies | Chicago boundaries for geo merging | | N/A |

Since the objective of this project is to compare probability of failure of business in time spans of one-year vs three years. We need to construct the duration. In the graph xx we see most of the business duration is concentrated in the less of 1000 days. Actually, 46% of the cases have duration over 3 years. On other hand, only 16% have a duration less than a year. Finally, we have business that duration between one year and two years are 38%.



The duration of the business also is variates trough in which part of the city the business is located, we can see that using a quintile distribution of duration we see that it’s concentrated in certain parts of the city having the biggest concentration in the yellow marked parts of the map.

