**HW 3 Outputs**

**Sam Yuxiao Sun**

**Question 1: Handlers Exploration**

The hhandler table contains registration forms for hazardous waste handlers. See the data dictionary here.

**(a) How many forms have been received by the EPA?**

2280396

**(b) How many facilities have registered? Hint: use the count(distinct ) function**

856529

**(c) How many forms were received per year over the last 5 years?**

(I am retrieving data for year 2011-2015)

|  |  |
| --- | --- |
| Year | Count |
| 2011; | 69795 |
| 2012; | 92413 |
| 2013; | 87378 |
| 2014; | 128099 |
| 2015; | 59757 |

**Question 2: Evaluations**

The cmecomp3 table contains a list of evaluations (inspections) of these handlers. See the data dictionary here.

**(a) How many evaluations are there?**

2061000

**(b) How many evaluations found violations?**

1478260

**(c) What proportion of evaluations found violations?**

0.717253760310529

**(d) Which five handler\_ids have been found in violation the most times? How many times? Also find these handlers' site names in the hhandlers table.**

|  |  |
| --- | --- |
| hhandler\_id | site name |
| ILD000714881 | KEYSTONE GROUP BARTONVILLE PLT |
| ILD000714881 | KEYSTONE STEEL & WIRE CO |
| ILD006296800 | MUSICK PLATING CO |
| ILD048843809 | CHEMETCO |
| ILD048843809 | CHEMETCO INC |
| KYD053348108 | SAFETY-KLEEN CORP. |
| KYD053348108 | SAFETY-KLEEN SYSTEMS INC |
| KYD053348108 | SAFETY-KLEEN SYSTEMS, INC |
| KYD053348108 | SAFETY-KLEEN SYSTEMS, INC. |
| KYD088438817 | L.W.D. INC. |
| KYD088438817 | LWD INC |
| KYD088438817 | LWD, INC. |
| KYD088438817 | LWD INC CALVERT CITY |

(some ids correspond to multiple names, either because they have multiple names, or because of recording)

**Question 3: Industries**

The North American Industry Classification System is a system used by federal agencies to classify a business according to its industry. The naics table contains this information as retrieved from here. Start by skimming this file.

**(a) How many different naics codes are there? How many six-digit industry classifications are there? How many two-digit classifications are there? These determine the sectors as described here.**

2016 total, 978 six-digit, 20 two-digit (or 19, since 99---- represents industries not specified).

**(b) The hnaics table contains naics codes for some handlers. How many handlers have naics codes? How many don't?**

438053 with codes, 418476 without codes.

**(c) Join the hnaics table with the naics table and use a GROUP BY to determine which the number of facilities in each sector. Which sector has the most hazardous-waste handlers? The least?**

|  |  |
| --- | --- |
| Sector | Handlers |
| Other Services (except Public Administration)  **MOST** | 73240 |
| Retail Trade | 69456 |
| Transportation and Warehousing | 26939 |
| Administrative and Support and Waste Management and Remediation Services | 25574 |
| Wholesale Trade | 22392 |
| Construction | 22086 |
| Utilities | 19397 |
| Agriculture, Forestry, Fishing and Hunting | 16911 |
| Health Care and Social Assistance | 13957 |
| Educational Services | 12813 |
| Professional, Scientific, and Technical Services | 11367 |
| Real Estate and Rental and Leasing | 7237 |
| Manufacturing | 6808 |
| Information | 6428 |
| Mining, Quarrying, and Oil and Gas Extraction | 6163 |
| Arts, Entertainment, and Recreation | 2971 |
| Accommodation and Food Services | 1227 |
| Finance and Insurance | 955 |
| Management of Companies and Enterprises **LEAST** | 426 |

(note sector 99 is left because there are no matches)

**(d) Create a temporary table called hsectors containing unique pairs of handler ids and sector descriptions.**

Table stored in memory

**(e) Join hsectors to cmecomp3, to determine for each sector, the number of handlers evaluated, the number of evaluations, the number of violations, and the proportion of evaluations finding violations. Which sector has the most violations? The highest proportion of evaluations finding violations?**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sector | Handlers | Evaluations | Violations | Violations% |
| Administrative and Support and Waste Management and Remediation Services | 8171 | 230420 | 151904 | 65.92% |
| Other Services (except Public Administration) | 33098 | 203775 | 141507 | 69.44% |
| Wholesale Trade | 11526 | 153741 | 104786 | 68.16% |
| Transportation and Warehousing | 10539 | 124468 | 82247 | 66.08% |
| Retail Trade | 17976 | 101011 | 66422 | 65.76% |
| Professional, Scientific, and Technical Services | 5188 | 72165 | 51377 | 71.19% |
| Construction | 6006 | 60316 | 44985 | 74.58% |
| Educational Services | 3893 | 43778 | 32779 | 74.88% |
| Manufacturing | 3670 | 49677 | 32578 | 65.58% |
| Agriculture, Forestry, Fishing and Hunting | 5371 | 45215 | 31898 | 70.55% |
| Health Care and Social Assistance | 3929 | 39112 | 28918 | 73.94% |
| Information | 1942 | 36310 | 24982 | 68.80% |
| Utilities | 4913 | 42135 | 24904 | 59.11% |
| Real Estate and Rental and Leasing | 2620 | 35119 | 19370 | 55.16% |
| Mining, Quarrying, and Oil and Gas Extraction | 2458 | 22911 | 14663 | 64.00% |
| Arts, Entertainment, and Recreation | 1193 | 11897 | 8072 | 67.85% |
| Finance and Insurance | 340 | 5552 | 2795 | 50.34% |
| Accommodation and Food Services | 396 | 2889 | 1822 | 63.07% |
| Management of Companies and Enterprises | 207 | 3055 | 1805 | 59.08% |