/\* DATA EXPLORATION \*/

--Missing values

|  |  |
| --- | --- |
| **Column name** | **Number of missing values** |
| location | 2166 |
| food | 9540 |
| ingredient | 17615 |
| species | 8626 |
| serotype | 15991 |
| status | 9408 |

--Data quality issues

1. Redundant categories describing the same products in food field. Species field name is misleading as it contains both species and chemical or other agents. To simplify the outcome two additional tables were created:

* food\_categories
* species\_categories

1. 3357 records contain 0 in ‘illnesses’ field, which doesn’t make sense for quantitative analysis.

--Number of records (not null) overall

15762

--Number of records (not null) per year

|  |  |
| --- | --- |
| **year** | **count** |
| 1998 | 1098 |
| 1999 | 1092 |
| 2000 | 1100 |
| 2001 | 1011 |
| 2002 | 1032 |
| 2003 | 851 |
| 2004 | 1097 |
| 2005 | 773 |
| 2006 | 1054 |
| 2007 | 904 |
| 2008 | 870 |
| 2009 | 577 |
| 2010 | 728 |
| 2011 | 677 |
| 2012 | 729 |
| 2013 | 697 |
| 2014 | 719 |
| 2015 | 753 |

--Number of records (not null) per food category

|  |  |
| --- | --- |
| **food\_category** | **count** |
| Beverages | 99 |
| Bread | 30 |
| Dairy | 410 |
| Eggs | 111 |
| Fish and Seafood | 1419 |
| Fruit | 145 |
| Meat | 2258 |
| Multiple Foods | 334 |
| Mushrooms | 32 |
| Nuts | 11 |
| Oil | 3 |
| Other | 743 |
| Rice and Groats | 317 |
| Sweets and Pastry | 180 |
| Unknown | 11 |
| Vegetables | 974 |
| Water | 74 |

--Number of records (not null) per species category

|  |  |
| --- | --- |
| **species\_category** | **count** |
| Bacteria | 4489 |
| Chemicals | 11 |
| Heavy metals | 8 |
| Other | 41 |
| Parasite | 81 |
| Toxin | 843 |
| Unknown | 14 |
| Virus | 4582 |

--Number of illnesses, hospitalizations and fatalities registered overall

|  |  |  |
| --- | --- | --- |
| **total\_illnesses** | **total\_hospitalizations** | **total\_fatalities** |
| 280967 | 10675 | 280 |

/\* Q1: HOW HAVE FOODBORNE DISEASE OUTBREAKS EVOLVED OVER TIME? \*/

--FBD statistics by month

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **month** | **number\_of\_records** | **total\_illnesses** | **total\_hospitalizations** | **total\_fatalities** |
| January | 1278 | 23901 | 577 | 22 |
| February | 1201 | 20886 | 634 | 21 |
| March | 1424 | 24752 | 985 | 16 |
| April | 1432 | 28672 | 895 | 14 |
| May | 1554 | 26236 | 854 | 23 |
| June | 1448 | 24362 | 1133 | 25 |
| July | 1252 | 21946 | 1373 | 57 |
| August | 1268 | 20797 | 1149 | 18 |
| September | 1018 | 17051 | 668 | 12 |
| October | 1128 | 20805 | 1052 | 44 |
| November | 1238 | 24725 | 809 | 17 |

--FBD statistics by year

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **year** | **number\_of\_records** | **total\_illnesses** | **total\_hospitalizations** | **total\_fatalities** |
| 1998 | 1098 | 22093 | 781 | 31 |
| 1999 | 1092 | 18580 | 409 | 6 |
| 2000 | 1100 | 19520 | 529 | 18 |
| 2001 | 1011 | 18495 | 487 | 10 |
| 2002 | 1032 | 17538 | 519 | 14 |
| 2003 | 851 | 16537 | 485 | 19 |
| 2004 | 1097 | 20754 | 477 | 19 |
| 2005 | 773 | 15402 | 420 | 5 |
| 2006 | 1054 | 23743 | 947 | 9 |
| 2007 | 904 | 15838 | 638 | 15 |
| 2008 | 870 | 15811 | 523 | 8 |
| 2009 | 577 | 11256 | 444 | 7 |
| 2010 | 728 | 10934 | 529 | 19 |
| 2011 | 677 | 10810 | 758 | 43 |
| 2012 | 729 | 11164 | 564 | 16 |
| 2013 | 697 | 10880 | 907 | 11 |
| 2014 | 719 | 9964 | 525 | 16 |
| 2015 | 753 | 11648 | 733 | 14 |

--Most common causes of FBD over the years

|  |  |
| --- | --- |
| **year** | **species** |
| 1998 | *Salmonella enterica* |
| 1999 | *Norovirus* genogroup I |
| 2000 | *Norovirus* genogroup I |
| 2001 | *Norovirus* genogroup I |
| 2002 | *Norovirus* genogroup I |
| 2003 | *Norovirus* genogroup I |
| 2004 | *Norovirus* genogroup I |
| 2005 | *Norovirus* genogroup I |
| 2006 | *Norovirus* genogroup I |
| 2007 | *Salmonella enterica* |
| 2008 | *Norovirus* genogroup I |
| 2009 | *Salmonella enterica* |
| 2010 | *Norovirus* genogroup II |
| 2011 | *Norovirus* genogroup II |
| 2012 | *Norovirus* genogroup II |
| 2013 | *Salmonella enterica* |
| 2014 | *Salmonella enterica* |
| 2015 | *Norovirus* genogroup II |

/\* Q2: WHICH FOODS AND LOCATIONS POSE THE HIGHEST RISK FOR FOODBORNE DISEASE CONTAMINATION? \*/

--Top 3 most dangerous foods

|  |  |
| --- | --- |
| **food\_category** | **count** |
| Meat | 2258 |
| Fish and Seafood | 1419 |
| Vegetables | 974 |

--Meat type most frequently causing FBD

|  |  |
| --- | --- |
| **meat\_type** | **count** |
| Poultry | 1048 |
| Beef | 638 |
| Pork | 281 |

--The most common locations and their FBD data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **location** | **report\_count** | **illnesses\_per\_report** | **hospitalizations\_percentage** | **deaths\_percentage** |
| Restaurant | 9099 | 12 | 3.45 | 0.05 |
| Private Home/Residence | 1507 | 12 | 11.71 | 0.40 |
| Catering Service | 895 | 29 | 1.16 | 0.01 |
| Fast Food Restaurant | 343 | 10 | 7.85 | 0.00 |
| Banquet Facility | 309 | 33 | 0.66 | 0.01 |
| School/College/University | 308 | 55 | 1.12 | 0.00 |
| Grocery Store | 252 | 13 | 6.73 | 0.28 |
| Nursing Home/Assisted Living Facility | 171 | 29 | 5.28 | 0.83 |
| Prison/Jail | 152 | 100 | 0.91 | 0.03 |
| Camp | 109 | 36 | 1.03 | 0.00 |
| Religious Facility | 95 | 30 | 5.40 | 0.00 |
| Office/Indoor Workplace | 83 | 41 | 1.38 | 0.00 |
| Fair/Festival | 72 | 21 | 5.39 | 0.00 |
| Hospital | 42 | 26 | 4.77 | 0.54 |
| Child Daycare | 26 | 20 | 6.98 | 0.00 |
| Restaurant Buffet | 7 | 8 | 3.28 | 0.00 |
| Farm/Dairy | 7 | 17 | 5.88 | 0.84 |
| Hotel/Motel | 1 | 7 | 0.00 | 0.00 |
| Ship/Boat | 1 | 3 | 0.00 | 0.00 |

/\* Q3: WHICH PATHOGENES PRESENT THE GREATEST THREAT TO PUBLIC HEALTH? \*/

--Number of illnesses, hospitalizations and deaths by species category

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **species\_category** | **total\_illnesses** | **illnesses\_rank** | **total\_hospitalizations** | **hospitalizations\_rank** | **total\_deaths** | **deaths\_rank** |
| Virus | 114180 | 1 | 1390 | 2 | 23 | 2 |
| Bacteria | 95368 | 2 | 8034 | 1 | 228 | 1 |
| Toxin | 3958 | 3 | 334 | 3 | 8 | 3 |
| Parasite | 2179 | 4 | 46 | 4 | 0 | 5 |
| Other | 205 | 5 | 13 | 5 | 0 | 5 |
| Unknown | 110 | 6 | 7 | 6 | 1 | 4 |
| Heavy metals | 102 | 7 | 1 | 8 | 0 | 5 |

--FBD caused by biological agents by genus (for viruses, bacteria and parasites)

|  |  |  |  |
| --- | --- | --- | --- |
| **genus** | **total\_illnesses** | **total\_hospitalizations** | **total\_deaths** |
| *Anisakis* | 14 | 0 | 0 |
| *Astrovirus* | 36 | 0 | 0 |
| *Bacillus* | 2243 | 22 | 0 |
| *Bacterium* | 1013 | 4 | 0 |
| *Brucella* | 14 | 6 | 0 |
| *Campylobacter* | 6203 | 232 | 1 |
| *Clostridium* | 20813 | 252 | 22 |
| *Cryptosporidium* | 520 | 11 | 0 |
| *Cyclospora* | 1187 | 18 | 0 |
| *Enterococcus* | 13 | 0 | 0 |
| *Escherichia* | 8275 | 1504 | 29 |
| *Giardia* | 343 | 1 | 0 |
| *Hepatitis* | 2322 | 237 | 5 |
| *Listeria* | 700 | 527 | 111 |
| *Norovirus* | 108773 | 1129 | 11 |
| *Parasite* | 21 | 3 | 0 |
| *Rotavirus* | 178 | 4 | 7 |
| *Salmonella* | 41260 | 4883 | 61 |
| *Sapovirus* | 54 | 0 | 0 |
| *Shigella* | 6956 | 263 | 1 |
| *Staphylococcus* | 6605 | 284 | 1 |
| *Streptococcus* | 98 | 4 | 0 |
| *Trichinella* | 94 | 13 | 0 |
| *Vibrio* | 1067 | 45 | 2 |
| *Virus* | 2817 | 20 | 0 |
| *Yersinia* | 108 | 8 | 0 |

--The deadliest cause of FBD

|  |  |
| --- | --- |
| **species** | **total\_deaths** |
| *Listeria monocytogenes* | 111 |