WSDOT DISCLOSURE REQUESTS

Week Two: Understanding Disclosure Datasets

By Angela Gonzalez-Curci for WSDOT as part of University of Washington/Open Data Literacy Internship

# Purpose

The purpose of this document is to document the key takeaways of the interviews I conducted over the course of week two and document the work I have done in cleaning and understanding the Public Record Disclosure Dataset.

# Analytics Workflow

Progress as of Week Two

## Data Dictionary

The purpose of the Data Dictionary is to define the meaning and structure of the columns pulled to create the PDR dataset I am working with.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Column Name | Business Term | Definition | Data Type | Data Range | Format | Scale | Null |
| Assigned | Public Disclosure Coordinator | First and Last Name of the Public Disclosure Coordinator Assigned to a Request. | Varchar |  | First Last |  | No |
| Assigned Group |  | The location the assigned Public Disclosure Coordinator works out of. | Varchar |  | FMS\_PDR\_ “Location Code” |  | No |
| Collision  DenialNY |  | Whether or not a requestor has signed a wavier denying that they attempt to use requested collision data to sue WSDOT. | Varchar |  | Y for Yes  N for No  NaN for not Collision |  | Yes |
| Collision Location |  | The location of a collision, which a data request has been filed for. | Varchar |  |  |  | Yes |
| Company |  | The organization, not the individual, filing a request. | Varchar |  |  |  | Yes |
| FullNameFML |  | The First, Middle (if recorded), and Last Name of the individual who filed the request. Note: Company and Full Name are not mutually exclusive. The individual can be from a requesting organization. | Varchar |  | First M. Last |  | Yes |
| Copy Size (1-4) | Media | The type of media a record is given to a requestor on. Up to four different type of media were sent to a requestor. | Varchar |  |  |  | Yes |
| CopyQuantityTotal |  | The total number of copies of all type of media sent to a requestor. Derived column from Copy Size 1-4. | Integer | 1-14958 |  | 1 | Yes |
| CopyTotalAmount |  | Total cost of copies of records request sent to a requestor. | Currency | $1.20-$4765.20 |  | 1.00 | Yes |
| DenialComments |  | Comments about a denied request added by the assigned coordinator | Varchar |  |  |  | Yes |
| DenialItems |  | Material in request which caused it the full request to be denied. | Varchar |  |  |  | Yes |
| ExemptionNumber (1-5) |  | Statue or court ruling cited as for reason for an exemption. Up to five different cited reasons for an exemption. Note: An exemption does not mean a full denial of a request, but rather that an original request was not given in full. | Varchar |  |  |  | Yes |
| Disposition |  | The final determination of the Public Disclosure Request. | Varchar |  |  |  | Yes |
| Disposition2 |  | Extra information on the final determination of the Public Disclosure Request not included in the Disposition Field | Varchar |  |  |  | Yes |
| ExtensionNumber |  | The length of time in days of an extension, if an extension was granted for a request | Integer | 5-90 |  | 1 | Yes |
| ExtensionComments |  | Comments about an extension added by the assigned coordinator. | Varchar |  |  |  | Yes |
| FinalCloseYN |  | Whether or not a Public Disclosure Request has officially been closed. | Varchar |  | Y for Yes |  | Yes |
| RequestedItems |  | The verbatim text of an original Public Disclosure Request. | Varchar |  |  |  | No |
| RequestTopic |  | The assigned public disclosure coordinator’s interpretation of an original Public Disclosure Request | Varchar |  |  |  | No |
| RID |  | The unique record identifier for a request. | Varchar |  | Two Variations:  RID-Year – String  RID-15-1234  -YearString  -151234 |  | No |
| State |  | The state the requestor is located in | Varchar |  | State Abbreviation  WA |  | Yes |
| Zip |  | The zip code the requestor is located in | Varchar |  | Two Variations:  Five Digit  12345  Five Digit – Four Digit  12345-6789 |  | Yes |

## Data Cleaning

I have been using Python to run an exploratory analysis on the PDR Dataset, and Jupiter Notebook to record every step of my process. Please see the attached Jupyter Notebook page for this documentation.

## Interview Questions

Managerial Questions

* What are particular to the types of records that you manage?
  + Particular types of requests?
  + Particular reasons for exemptions?
* How are your public records are being used? (By stakeholders or institutions?)
* Can you think of any important stakeholders that use data or information from your organization as part of their own work - such as through a data sharing agreement or open data portal?
* What challenges might you foresee in publishing public records proactively as open data?
  + What challenges in do you see in choosing public records to be proactively published?

Coordinator Questions

* Work with me through the record request process:
  + How do you know if there is an exemption for a request or not?
    1. (Is it based on institutional knowledge, the text itself, or something else?)
  + How do you deal with unclear requests?
    1. Clarification
    2. Is it based on institutional knowledge, the text itself, or something else?
* In your experience what are the most frequently requested records?

Were a request came in, determined coordinator

* + Are exemptions frequent?
  + Are exemptions circumstantial?
* What are the tools you use/used ?

How do/did they help you?

* What types of tools and resources do you wish you/ you’d had?
  + In handling record requests process?
  + In assessing the nature of a record request?