**Data Science Team Innovia**

# Data Sources:

## How to Use:

Each dataset has the same sections, below is a description of each section and how they can be used.

* **How to get the data:** Step by step on how to get the most up-to-date version of the
* **Column Descriptions:** The names of columns and what they represent, sometimes just notes
* **Filtering Columns:** Columns that can be useful for cleaning the data from unhelpful rows
* **Linking Columns:** Identifying columns that can be used to match individuals across datasets
* **Wealth ID Columns:** Useful columns for determining the wealth of an individual
* **Age ID Columns:** Useful columns for determining the age of an individual
* **Donor ID Columns:** Useful columns for determining the inclination to donate of individual

## Political Donations Data:

### How to get the data:

1. [Click Here](https://www.pdc.wa.gov/browse/open-data/contributions-candidates-and-political-committees) (https://www.pdc.wa.gov/browse/open-data/contributions-candidates-and-political-committees)
2. Press “Explore this dataset at data.wa.gov” halfway down the page
3. Press the “Export” tab on the right hand side
4. Press “CSV” and follow any browser prompts to finish the download

### Column Descriptions:

* **code** - One of the following types of entities (Anonymous, Business, PAC, Individual, Caucus, Other, Bona Fide Party, Self Pers. Fund, Minor Party, Union)
* **report\_number** - multiple contributions or corrections will correspond to the same report number
* **party** - One of these based on donation receiver (REPUBLICAN, DEMOCRAT, NON PARTISAN, NA)
* **election\_year -** approximate year to which the donation corresponds to, check website description
* **contributor\_name** - contains capitalized full name of the donor, can also be company name
* **contributor\_address -** might not be accurate
* **contributor\_city**
* **contributor\_state**
* **contributor\_zip -** might not be accurate
* **contributor\_occupation -** only applies to individuals who donated more than $100

### Filtering Columns:

* **code** - may need to only accept “Individuals” type
* **report\_number** - need to combine all rows with the same report number

### Linking Columns:

* **contributor\_name**
* **contributor\_address**
* **contributor\_city**
* **contributor\_state**
* **contributor\_zip**

### Wealth ID Columns:

* **party -** richer people tend to vote more republican
* **contributor\_occupation -** can find median income of position

### Age ID Columns:

* **party -** older folks tend to be more republican
* **election\_year** - the more older donators could indicate older age
* **contributor\_occupation** - since some say “RETIRED”

### Donor ID Columns:

* **party -** possibly a correlation for democrats being more donor friendly?
* **election\_year** - could get a recent donator feature
* **contributor\_occupation** - since some say “RETIRED”

## Small Business Association (SBA) Data:

### How to get the data:

1. [Click Here](https://web.sba.gov/pro-net/search/dsp_dsbs.cfm) (https://web.sba.gov/pro-net/search/dsp\_dsbs.cfm)
2. Select the state for which to gather data (Probably WA or ID)
3. Collect two datasets one for businesses upto 5 employees and one with revenue over $500000
4. Get the columns matching the below description by selecting the columns in the appropriate dialog at the end
5. Select mail-merge format, and delimeter as a comma

### Column Descriptions:

* **Company Name**
* **Contact Name**
* **Address, City, City, Zip**
* **(implied) Number of employees all less that**
* **Accepts Credit Cards**
* **County Code**
* **Email Address**
* **Year Established**
* **WWW Page URL**
* **Key words**
* **Special Equipment**

### Filtering Columns:

* **County Code** - may want to limit results to eastern part of the state

### Linking Columns:

* **Company Name -** if any organization matches the general location
* **Contact Name -** usually the owner can cross reference with other names along with address
* **Address -** Can match business or individuallocations
* **Email Address**

### Wealth ID Columns:

* **(implied from query) revenue -** based on dataset can filter how much money for how many people the business is earning
* **# of employees -** the more employees the more the profits usually
* **Year established -** the longer the business works the more wealth the individual is usually

### Age ID Columns:

* **Year established -** the longer the business works the more years the owner lived
* **Accepts Credit Cards -** if accepts this payment type the person is generally younger
* **Company Website -** if has something already person is younger

### Donor ID Columns:

* **(implied from query) # of employee’s ratio to revenue -** all the same
* **#**