**Assignment 2: Tableau basics and data exploration with histograms and bar/column charts**

**Objective:**

Explore Tableau’s interface, data organization features such as field hierarchies, and simple calculated fields while exploring data using histogram and bar/column charts.

**Dataset:**

NYC voter registration Feb 2017 data (county to election district-level data).

**Deliverables:**

1. Your Tableau workbook
2. Word doc answering the questions posed below.

**Data Cleaning and Preparation Instructions**

1. Clean the Dataset using Excel

* Remove blank rows and unnecessary columns such as totals since you can easily calculate the totals.

1. Connect and explore the data structure in Tableau (data types, metadata, and dimensions vs. measures).

* Connect to the voter registration dataset in Tableau.
* Move the following columns to Dimensions: County, Council District, Assembly District (AD), Election District (ED).
* Move all voter registration count columns (e.g., Active, Inactive Count for each party) to Measures.

1. Create a Hierarchy:

* Under Dimensions, create a hierarchy named ElectionGeoLevels.
* Add County, Council District, Assembly District, and Election District to the hierarchy, in that order.

1. Create the following Calculated Fields:

* In the Data Pane, click the Drop-Down Menu and select Create Calculated Field for each of the following:

1. ConLeaning\_ACTIVE: [CON Active] + [REF Active] + [REP Active]
2. LibLeaning\_ACTIVE: [DEM Active] + [WEP Active] + [WOR Active]
3. Other\_ACTIVE: [BLANK Active] + [GRE Active] + [OTH Active] + [IND Active]
4. Total\_ACTIVE: Add above three fields
5. ConLeaning\_INACTIVE: [CON Inactive] + [REF Inactive] + [REP Inactive]
6. LibLeaning\_INACTIVE: [DEM Inactive] + [WEP Inactive] + [WOR Inactive]
7. Other\_INACTIVE: [BLANK Inactive] + [GRE Inactive] + [OTH Inactive] + [IND Inactive]
8. Total\_INACTIVE: Add above three fields

* Quickly verify the results:

1. Drag these calculated fields into the Rows or Columns shelf to verify the calculations by comparing results against the original data columns.

**Visualization Tasks:**

1. Bar/Column clustered vs. stacked charts

* Create a new worksheet called **‘ACTIVE**\_**VOTERS’**
* Create a stacked column chart here showing active registered voters by County from each of the newly calculated fields (plus the independent group). Since you are creating a stacked column chart using multiple columns (rather than multiple values in a single column), you will need to:

1. Move ‘Measure Names’ from Dimensions to Filters and select the columns (i.e. calculated fields you just created) to include in the visualization
2. Move the ‘Measure Values’ from the Measures to Rows shelf
3. Move County to the Columns Shelf
4. Copy the ‘Measure Names’ from Filters to Color card in the Marks area.
5. Edit and assign distinct colors to each party group [Choose the party colors used commonly to represent them]

* In a new worksheet called ‘INACTIVE VOTERS’, repeat previous steps 1-5 to create a clustered bar chart

1. Copy the ‘Measure Names’ to the Columns shelf also to transform the stacked column chart to clustered column chart.

* In a new worksheet “Active\_Inactive\_County’, follow the similar procedures to add fields for total active and inactive voters to compare voter activity at the county level.

1. Feel free to experiment with additional election hierarchy fields such as Districts, Assembly District, and Election District.

* Experiment with adding variables to different cards in the marks area.
* Be sure to include informative chart titles, axis labels, and data labels.

1. Histogram

* Storyline: How does voter engagement vary across the election districts? Highlight areas that might need more voter engagement initiatives or outreach programs.
* Create a new worksheet called **histo\_ACTIVE** and now create a histogram of Total\_ACTIVE voters by Election District using following steps

1. Choose Total\_ACTIVE field à Right-click à Create à Bins à Specify 50 for Size of Bins à Okay à Confirm Binned measure is added to Dimensions area.
2. Copy the Binned Total\_ACTIVE dimension field to the Columns shelf.
3. Copy the # of records (Count) measure to the Rows Shelf
4. Add and Show County filter
5. Add a label to display the count of districts in each bin.

* Repeat previous step for creating histogram for Total\_INACTIVE voters by Election District
* Be sure to include informative chart titles, axis labels, and data labels.

1. **Analysis and Reflection**

* Bar/column charts

1. Move the pills around in the columns shelf in both charts. As your visuals change, do you prefer one specific layout over the other? Why?
2. What questions are you able to answer using these charts?
3. Are you able to answer these questions easily with one layout vs. other? Why so?

* Histogram

1. What is the most common range of active/inactive voters across districts?
2. Are there outlier districts with significantly high or low active/inactive voter counts?
3. Is the distribution skewed towards any one end? How does the distribution change when filtering for specific counties?

* Things to look at: Is there a narrow/wide peak, single/multiple peaks, wide flat distribution, etc.
* What actions can campaign officials take based on these visuals?

**Submission Requirements**

1. Save your Tableau workbook as a packaged file (.twbx)
2. Rename these files with your first and last name and submit the .twbx and word file by the due date.