Table of Contents

Los Angeles Arrest Data 2010-2019: SAP Analysis

*Contributors: Gia DiGiacobbe, Bill Evashwick,   
Gunzal Gaji, Xiomara Iraheta, Raymond Tsang,*

**Section 1:** Data Preparation: Create Dataset

**Section 2:** Create Model**:** ERD Model

**Section 3:** Create Story: GeoMap

**Section 4:** Create Chart: Bar + Pie Chart

**Section 5:** Time Series

**Section 6:** Regression

**Section 1: Create a Data Set**

**Objective 1:**The objective of this exercise is to acquire and model Los Angeles Arrest Data for analysis in SAP Analytics Cloud.

*Data Set:* [***Arrest\_Data\_from\_2010\_to\_2019(10).xlsx***](https://github.com/BUS-5100-Group-2)

1. Acquire the Data

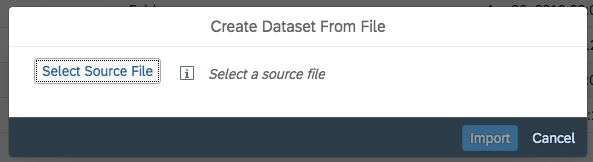
To use a data file in SAP Analytics Cloud (SAC) predictive scenarios, the file must be acquired and saved as an SAC dataset.

1. Open a web browser and go to <https://github.com/BUS-5100-Group-2>
2. Download the file ***Arrest\_Data\_from\_2010\_to\_2019(10)*** to your computer from **GitHub→ Downloads → personal file →** *Arrest\_Data\_from\_2010\_to\_2019(10)*

2. Create a dataset

Log onto SAP Analytics Cloud (SAC): <https://highereducation.us10.sapanalytics.cloud/>

1. Create → Data Set
2. Select the data uploaded from a file
3. Select Source File



1. Select ***Arrest\_Data\_from\_2010\_to\_2019(10)***
   1. Use the first row as column headers should be selected by default
2. Import
3. f. Save to your folder for the location
4. Name the file “Los Angeles Arrest Data”
5. Click **OK**
6. **Save** your data set

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Section 2: CREATE MODEL**

1. Go to Create → Model

2. Choose Import a file from your computer

3. Then, click **Select Source FiIe** Select the excel file *Arrest\_Data\_from\_2010\_to\_2019(10).xls* and open it.

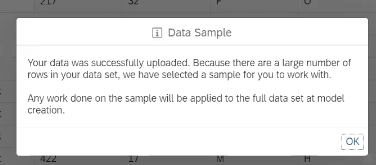
4. **Use the first row as column headers** should be selected as default. Click **Import**.

5. Import. This is a very large file so it might take some time, more than a minute.

6. Click on the data file name created in the import.

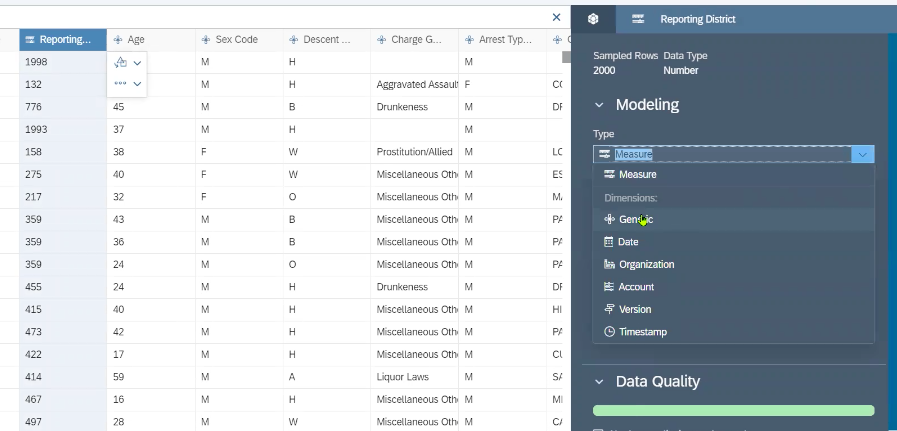
7. You will receive a message that only a sample of the data will be shown (see figure below).

Click **OK** to acknowledge the message. **NOTE:** You need to save the model when you do each step as not to lose any progress.



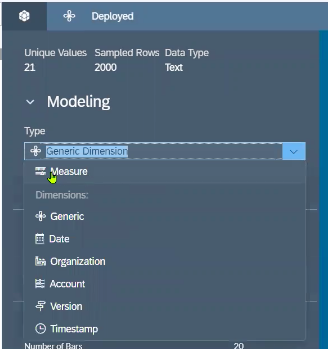
8. Select the column, *Reporting district* and make the following change:

* Change *Reporting District* data type to ***Generic Dimension***



9. Select the column, *Deployed* and make the following change:

* Change *Deployed* data type to ***Measure***



10. Select the column, *Counter* and make the following change:

* Change *Counter* data type to ***Measure***

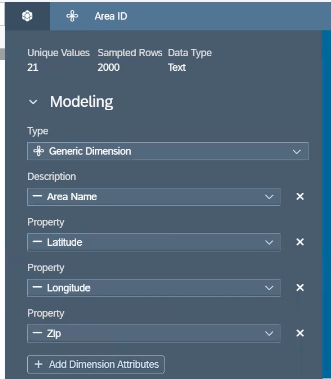
11. Select the column, *Area ID* and add attributes. Add the following Dimension Attributes to *Area ID:*

(1) For Description add ***Area Name*** *from column list* to link the ***Area ID*** and ***Area Name*.**

(2) Add Property ***Longitude***

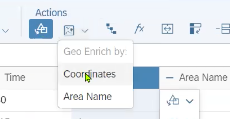
(3) Add Property ***Latitude***

(4) Add Property ***Zip***



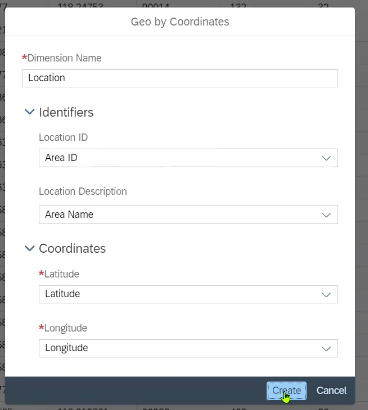
12. Add a geo enrichment based on latitude and longitude by Area ID.

In the Actions Ribbon, select Geo Enrichment → Coordinates



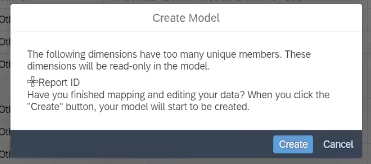
The following window will appear - select the following

* **Location ID**, select *Area ID.*
* **Location Description**, select *Area Name.*
* The Coordinates should automatically detect the dimensions *Latitude* and *Longitude.*
* Select **CREATE**



13. In the Builder menu, select **Create Model**

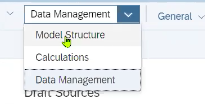
14. Once your data finishes validating, you will receive the following message window → select **Create**



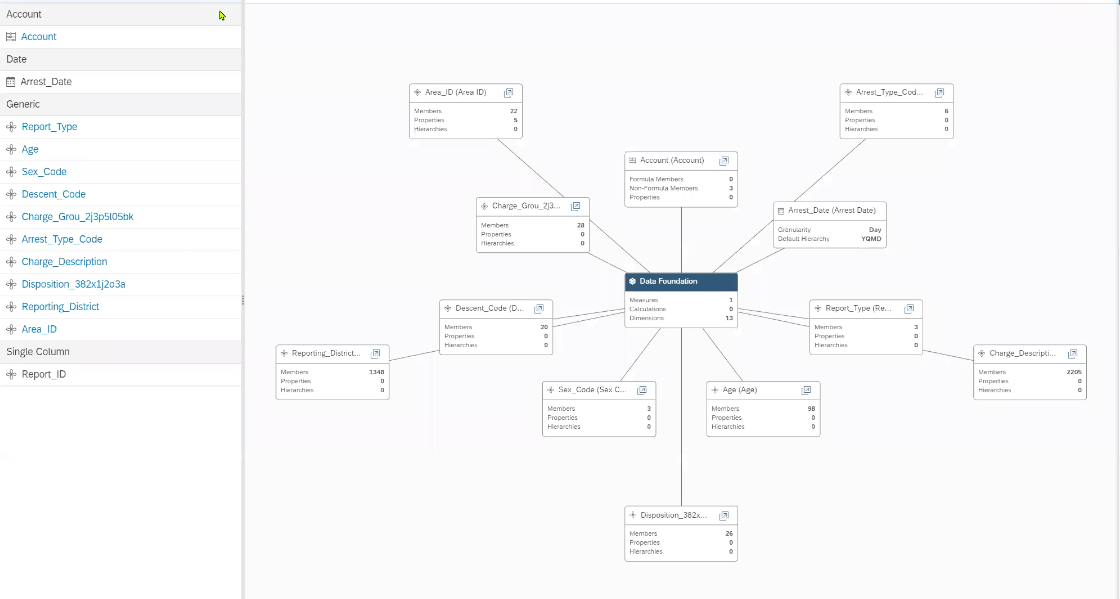
15. **SAVE** your model to your personal folder

* Select CalStateLA folder → your personal folder
* Name it *Arrest\_Data\_from\_2010\_to\_2019 (10).csv*
* Click **OK**

16. In the top left of the screen, select the **Model** Tab drop down menu, and change to *Model Structure*



17. The following Model Data Foundation View should appear

****

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

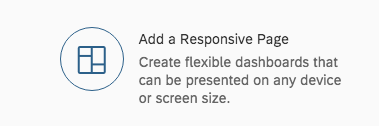
**Section 3: Create GeoMap**

**Create A Story (GEOMAP)**

*Objective:* Using the model created with our data set, we will explore various visualization of Arrest Data for Los Angeles; specifically, the correlation of crime types and locations.

1. Click the Menu table → Create → Story

* Select **Add Responsive Page**

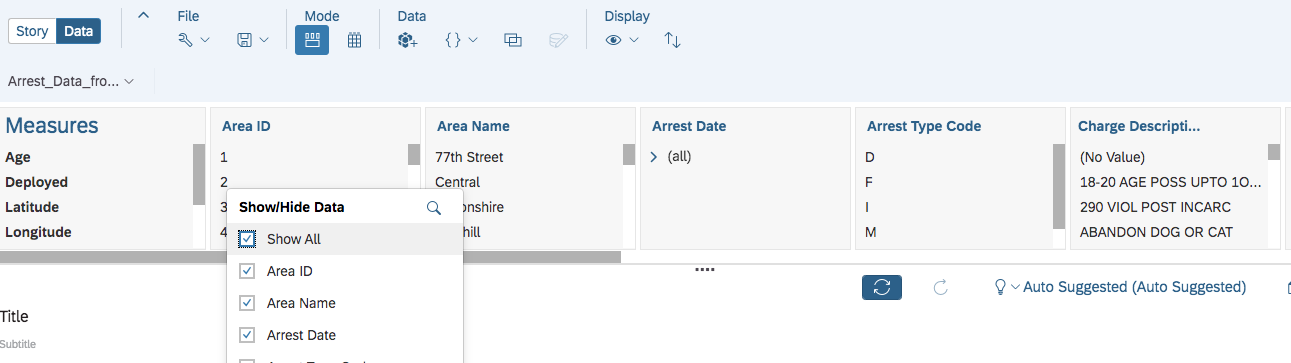


2. On the top left, toggle to **Data** view

* Choose to add *Data from existing dataset or model* (see figure below

Choose Cal State LA folder → your folder → select the model you created *“Arrest\_Data\_from\_2010\_to\_2019 (10)”*

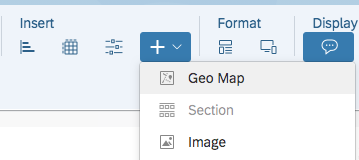
3. Click on **Show Dimensions** and select *Show All*



4. Save your story as “LA Crime Rate Story” or a name of your choosing.

5. Toggle back to **Story** tab.

Insert a Geo Map by selecting **Insert +** on the ribbon and choose***GeoMap***

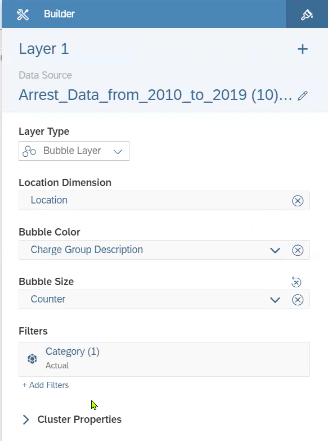


6. In the Designer Builder view choose “**Base Layer”** and select *OpenStreetMap* as the style of the background for visualization.

7 Under **Layer Type**, select *Bubble Layer.*

8. Then, under **Map Layers**, select *+ Add Layer*

* Under **Location Dimension**, select *Location*
* Under **Bubble Color**, select *Charge Group Description* listed under Dimensions
* Under **Bubble Size**, select *Counter* under Measures



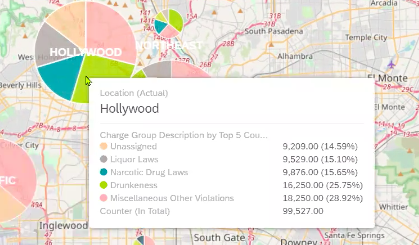
(If you want to adjust the visuals, select the drop-down menu under Bubble Color and select the color scheme of your choosing)

9 Toggle from the **Builder** view to *Styling*

* Under Size, select 16



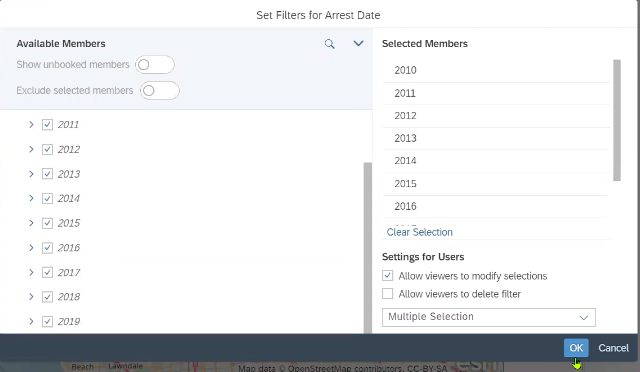
(You can use the zoom function to take a closer look at certain areas)



10. Under **Filters**, add a Category *cluster properties* from the drop down menu

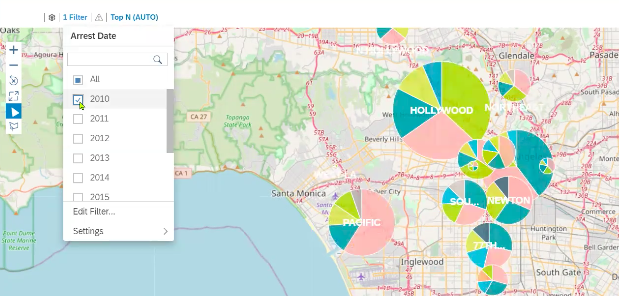
The following window will appear

1. Uncheck “*All*” and select all the individual years (2010-2019) → click **OK**

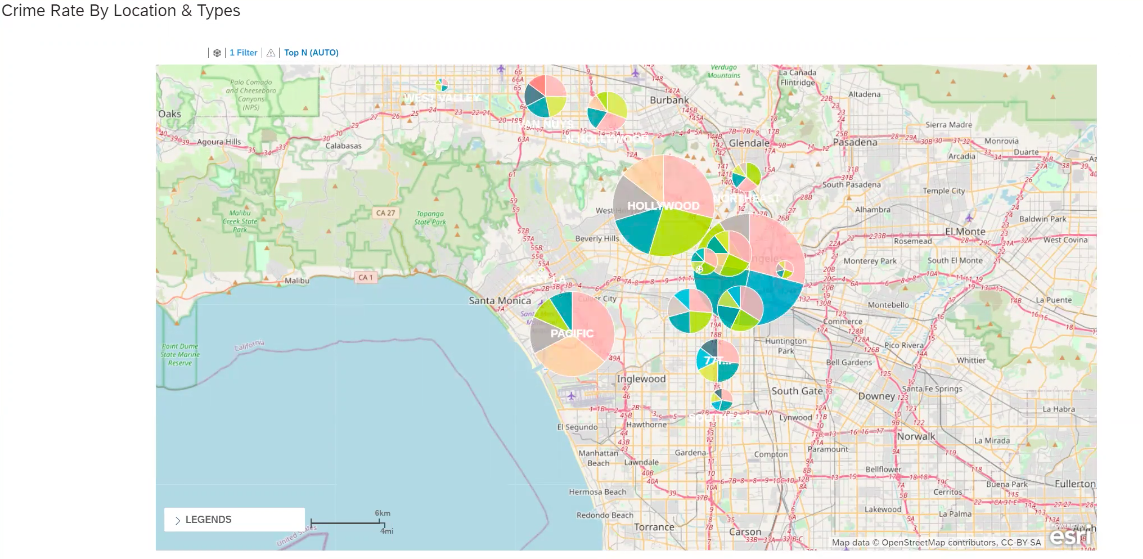


11. Once back to the Story view, click on the **filter** option at the top of the map

* From here you can select specific years to show the data pertaining to that time period



12. Add the Title “Crime Rate by Location & Types” to the Geomap.



13. **SAVE** your story.

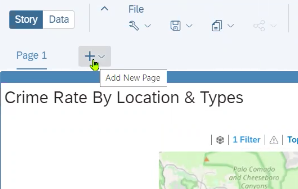
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Section 4: BAR/COLUMN + PIE CHART**

*Objective:* Utilize different chart types (Bar/Column + Pie) to observe relationships between frequency of arrest type and major metro areas within Los Angeles. It allows us to look at / discover trends across the decade spanning from 2010-2019 and predict what arrests might be most prevalent in the future. .

1. Add a second Responsive Page to the Story (Use the “+” next to Page 1)

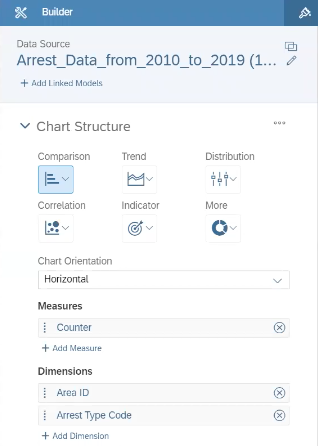
* Under + Select “Add Responsive Page”



2. In the new page, select Insert Chart and select the following features:

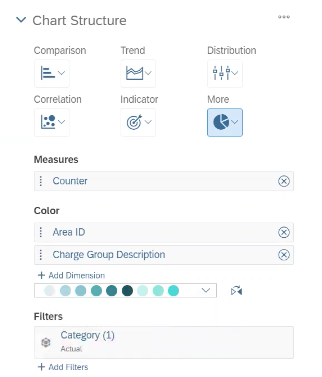
1. In the Builder view, under **Chart Structure** select **Comparison** choose ***Bar/Colum***
2. Under **Chart Orientation**, select ***Horizontal***
3. Under the **Measures** drop down menu → under **Account** select ***Counter***

3. Under **Dimensions**, add the following ***Area ID*** & ***Arrest Type Code***



4. Click over to the right lane of the page → select Insert Chart → More → **Pie Chart**

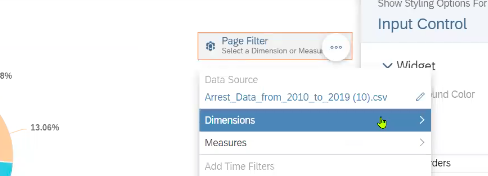
1. Under **Measures**, select “***Counter***”
2. Under **Color**, select ***Area ID*** & ***Charge Group Description***
   1. Select the color scheme you like that best displays the data



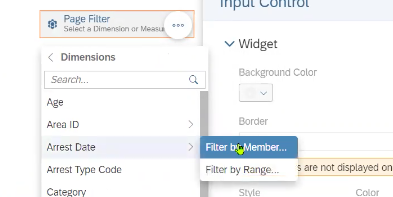
5. Toggle over to the **Styling** function

6. Under **Legend**, set the Placement to *below chart*

7. Select the right side of the page, then under the Insert Ribbon select **Input Control**

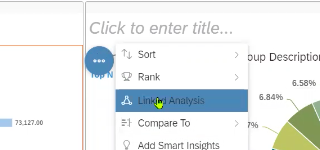


8. Click the dropdown menu under **Dimension** to populate this window:

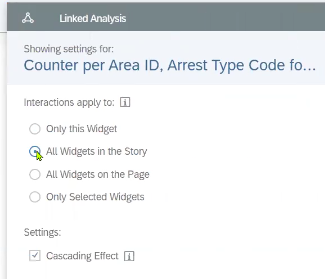


1. Select **Arrest Date** → Filter by **Member** → select *All Members* → click **OK**
   1. Select the page filter and drag it to the left side of the responsive page under the Bar/Colum Chart
   2. Expand the view

9. Click the More (“...”) option on the left side of the page, and select **Linked Analysis**

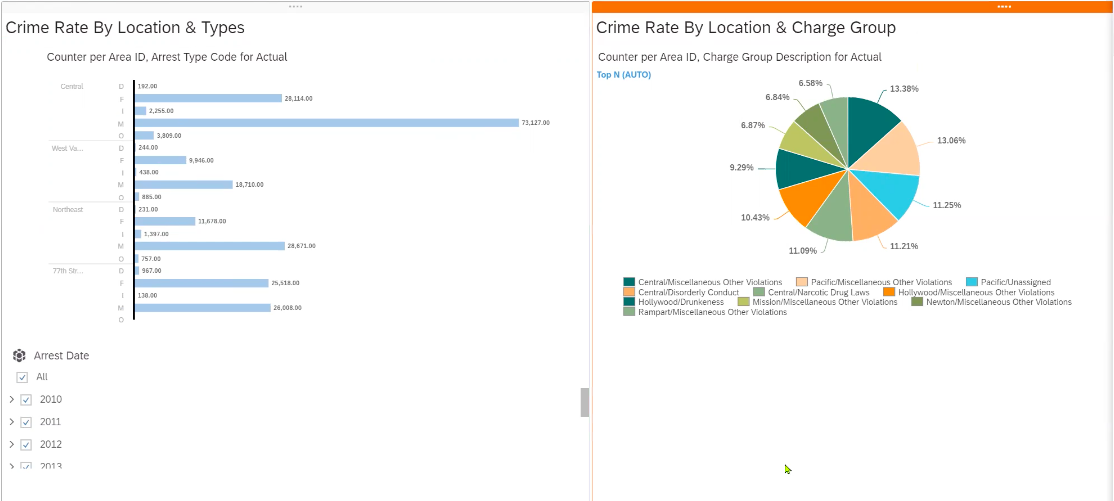


* Select **All Widgets in the Story** → click Apply



10. Add Titles to each chart

1. Add the title “Crime Rate by Location & Types” to the **Bar/Colum Chart** on the left side of the page
2. Add the title “Crime Rate by Location & Charge Group” to the **Pie Chart** on the right side of the page
3. You will see your two charts side-by-side (see figure below)

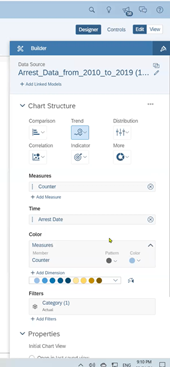


\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

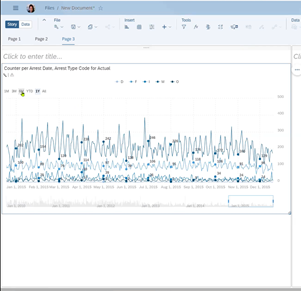
**Section 5: Time Series**

*Objective:* This activity provides an opportunity to explore some of the time series analysis functions in SAP Analytics Cloud in the context of Los Angeles Crime Data

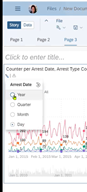
1. Add a new page
2. Hover your cursor next to “Page 2” at the upper left of your screen until the “+” sign appears
3. Click the +
4. Choose “Responsive”
5. Insert your time series
6. Under insert, click the chart icon
7. The builder will appear on the right of the screen
8. Settings
9. Trend: Time Series
10. Measures: Counter
11. Time: Arrest Date
12. Click “+ Add Dimension”
13. Choose “Arrest Type Code”
14. You may choose any color scheme of your liking
15. See screenshot reflecting these settings:



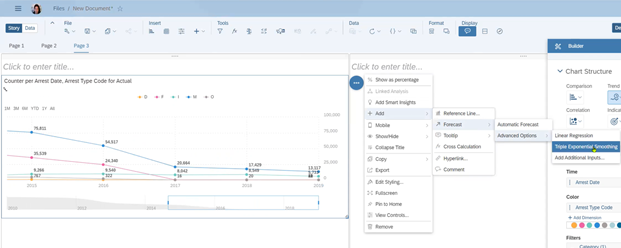
1. Your chart will now appear in the story panel on the left side of your screen
2. You may stretch the chart out by clicking and dragging the lower right corner
3. See screenshot of what your chart may look like



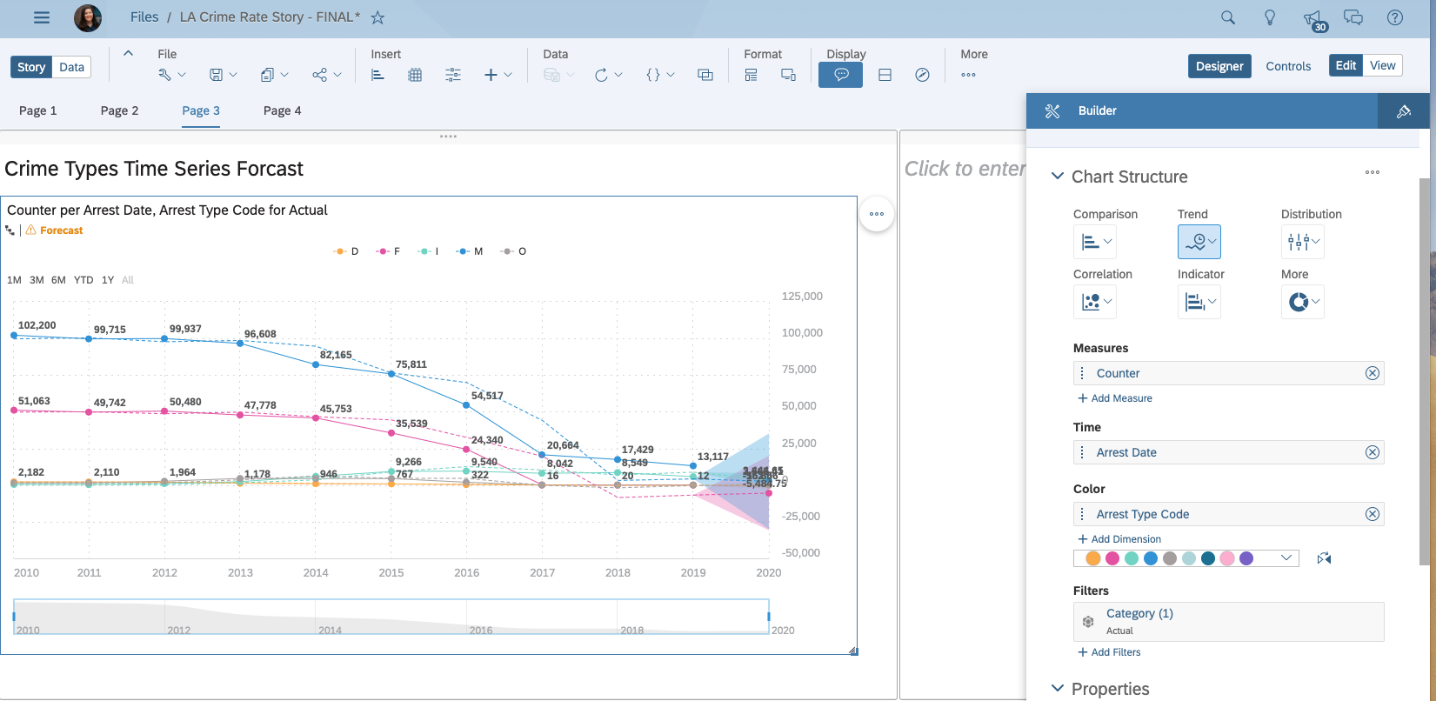
1. Click “Set Drill”
2. Choose “Year”



1. Add a Forecast
2. Click the three dots in the upper right corner of your chart
3. + Add -> Forecast -> Advanced Options -> Triple Exponential Smoothing
4. See screenshot, below, for the pathway



1. You can drag your slider to show different time periods, including your forecast
2. See screenshot, below, for an example of what your table will look like with the slider showing 2010 through the forecast



1. Give your story a name
2. Save your story into your folder

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

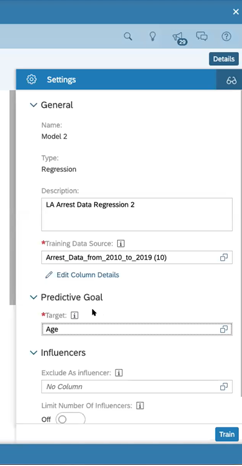
**Section 6: Regression**

*Objective:* The objective of this exercise is to create a regression model to help predict (forecast) the age of perpetrators of crime in Los Angeles.

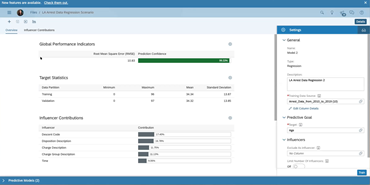
1. Acquire the Data

To use a data file in SAP Analytics Cloud (SAC) predictive scenarios, the file must be acquired and saved as an SAC dataset.

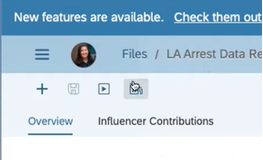
1. Open a web browser and go to SAP Analytics Cloud (SAC): <https://highereducation.us10.sapanalytics.cloud/>
2. Log-in
3. Download the file ***Arrest\_Data\_from\_2010\_to\_2019(10)*** from GitHub: https://github.com/BUS-5100-Group-2 to your computer from ***Arrest\_Data\_from\_2010\_to\_2019(10)***
4. Create a dataset
5. Create -> Data Set
6. Select the data uploaded from a file
7. Select Source File
8. Select ***Arrest\_Data\_from\_2010\_to\_2019(10)***
9. Use the first row as column headers should be selected by default
10. Import
11. Save to your folder for the location
12. Name the file “Los Angeles Arrest Data”
13. Click OK
14. Save your data set
15. Create the Regression Analysis Model
16. Create a Predictive Scenario
17. Create -> Predictive Scenario
18. Select Regression
19. Choose your folder to save your regression model and name it “Los Angeles Arrest Data Regression Scenario”
20. Select the dataset from “Los Angeles Arrest Data” in Settings, that you created in the previous step, as your Training Data Source
21. Under description title “LA Arrest Date Regression””
22. Predictive Goal -> *Target:* Age
23. *Influencers:* do not exclude any influencers
24. See screenshot for settings



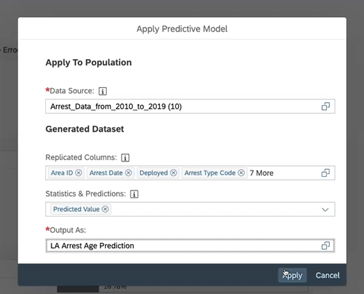
1. Click OK
2. Click Train, which will partition the data into a training set and a validation set
3. Examine the results
4. You will see two tabs, Overview and Influencer Contributions
5. See screenshot for view



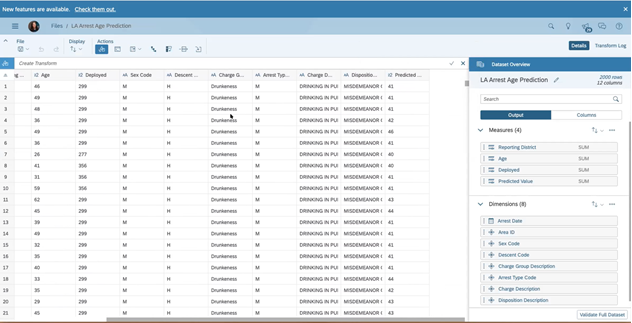
1. Save
2. You may choose to repeat this process but choose to exclude certain influencers to see how it affects your model and compare between different models. We have chosen to stick with this model because it has high accuracy and shows the contributions of many possible influencers
3. To check the accuracy of your model, Apply Predictive Model
4. Click on Apply Predictive Model
5. See screenshot, below to know where to click (grey highlighted button)



1. The settings are listed, below, and shown in the next screenshot
2. *Data Source:* Use your original data source to compare the predicted ages from the model to the actual values
3. *Replicated Columns:* are the columns that will be included in the output of the model. We chose Area ID, Arrest Date, Deployed, Arrest Type Code, Age, Reporting District, Descent Code, Disposition Description, Sex Code, Charge Group Description, Charge Description
4. Click OK
5. Statistics and Predictions: Predicted Values
6. Output as: “LA Arrest Age Prediction” and save to your own folder.



1. Select Apply
2. To see the results, open a new tab of SAC. Browse -> Files and select “Age Predictions” from your folder.
3. The predictions will be in the last column of the data set.
4. See screenshot, below
5. You can look at the column “Age” compared to the predicted values



**References:**

Original Data Link: data.lacity.org

Story Location: https://higher-education.us10.sapanalytics.cloud/

Cal State LA folder >Rtsang>Final Project Group 2

Github Link: <https://github.com/BUS-5100-Group-2/LAArrestData>