**ADTA-5250: Assignment 2: Tableau Prep**

**Part 1:**

Dataset selected: **Police\_Arrests.csv.**

Dataset structure: 65 columns; 101457 rows.

I have chosen this dataset since it contains data about Dallas crime which is common and reported on daily basis. I would love to show few insights about crime rate, severity zones and frequency using the learnings from Large Data Visualization.

**Data Dictionary**

Below table describes about data dictionary,

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Field Name** | **Description** | **Data Type** | **Data Format** | **Dimension/Measure** | **Example** |
| IncidentNum | Incident number or identifier | Text/String | Alphanumeric | Dimension | 053735-2023 |
| ArrestYr | Year of the arrest | Integer | YYYY | Dimension | 2023 |
| ArrestNumber | Arrest number or identifier | Text/String | Alphanumeric | Dimension | 23-008152 |
| ArArrestDate | Arrest date | Date | YYYY-MM-DD | Dimension | 30-Mar-23 |
| ArArrestTime | Arrest time | Time | HH:MM:SS | Dimension | 20:58 |
| ArBkDate | Booking date | Date | YYYY-MM-DD | Dimension | 31-Mar-23 |
| ArLAddress | Arrestee's street address | Text/String | Alphanumeric | Dimension | 1633 N COCKRELL HILL RD |
| ArLApt | Arrestee's apartment number | Text/String | Alphanumeric | Dimension | 123 |
| ArLZip | Arrestee's ZIP code | Text/String | Alphanumeric | Dimension | 75211 |
| ArLCity | Arrestee's city | Text/String | Alphanumeric | Dimension | DALLAS |
| ArState | Arrestee's state | Text/String | Alphanumeric | Dimension | TX |
| ArLCounty | Arrestee's county | Text/String | Alphanumeric | Dimension | DALLAS |
| ArLRA | Arrestee's reporting area | Text/String | Alphanumeric | Dimension | 4077 |
| ArLBeat | Arrestee's beat or precinct | Text/String | Alphanumeric | Dimension | 416 |
| ArLDistrict | Arrestee's district | Text/String | Alphanumeric | Dimension | 3 |
| ArLSector | Arrestee's sector or sector code | Text/String | Alphanumeric | Dimension | 410 |
| ArADOW | Arrestee's area of disposition | Text/String | Alphanumeric | Dimension | Thu |
| ArPremises | Location or premises of the arrest | Text/String | Alphanumeric | Dimension | Highway, Street, Alley ETC |
| CFS\_Number | Case or incident number (if applicable) | Text/String | Alphanumeric | Dimension | 23-0611304 |
| ArOfcr1 | Officer involved in the arrest (1st officer) | Text/String | Alphanumeric | Dimension | 10837 |
| ArOfcr2 | Officer involved in the arrest (2nd officer) | Text/String | Alphanumeric | Dimension | 22435 |
| Transport1 | Transportation details (1st transport) | Text/String | Alphanumeric | Dimension | 10837 |
| Transport2 | Transportation details (2nd transport) | Text/String | Alphanumeric | Dimension | 22435 |
| Transport3 | Transportation details (3rd transport) | Text/String | Alphanumeric | Dimension | 12345 |
| Search1 | Details of any search conducted | Text/String | Alphanumeric | Dimension | 54321 |
| ArAction | Action taken during the arrest | Text/String | Alphanumeric | Dimension | Arrested - Lew Sterrett |
| ArWeapon | Weapon involved (if applicable) | Text/String | Alphanumeric | Dimension | Unarmed |
| ArOSR | Arrestee's offense or reason for arrest | Text/String | Alphanumeric | Dimension | Public disturbance |
| ArResisted | Whether the arrestee resisted arrest | Text/String | Alphanumeric | Dimension | No |
| ArCurrLoc | Current location of the arrestee | Text/String | Alphanumeric | Dimension | LS |
| ArCond | Conditions surrounding the arrest | Text/String | Alphanumeric | Dimension | Calm |
| ArMedFlag | Medical flag (if applicable) | Text/String | Alphanumeric | Dimension | Yes |
| ArMedLoc | Medical location (if applicable) | Text/String | Alphanumeric | Dimension | DALLAS |
| ArOpComm | Operational comments or notes | Text/String | Alphanumeric | Dimension | Arrested |
| ArresteeName | Name of the arrestee | Text/String | Alphanumeric | Dimension | ESCALANTE ZAMORA, MODESTO |
| NickName | Arrestee's nickname (if available) | Text/String | Alphanumeric | Dimension | P J |
| AliasName | Alias or alternate name of the arrestee | Text/String | Alphanumeric | Dimension | Danny |
| BirthPlace | Place of birth of the arrestee | Text/String | Alphanumeric | Dimension | MEXICO |
| Age | Age of the arrestee | Integer | Years | Measure | 50 |
| AgeAtArrestTime | Age of the arrestee at the time of arrest | Integer | Years | Measure | 50 |
| HAddress | Home address of the arrestee | Text/String | Alphanumeric | Dimension | 822 SUNSET AVE |
| HApt | Home apartment number of the arrestee | Text/String | Alphanumeric | Dimension | 345 |
| HZIP | Home ZIP code of the arrestee | Text/String | Alphanumeric | Dimension | 75208 |
| HCity | Home city of the arrestee | Text/String | Alphanumeric | Dimension | DALLAS |
| HState | Home state of the arrestee | Text/String | Alphanumeric | Dimension | TX |
| HRA | Home reporting area of the arrestee | Text/String | Alphanumeric | Dimension | 4118 |
| HBeat | Home beat or precinct of the arrestee | Text/String | Alphanumeric | Dimension | 417 |
| Height | Height of the arrestee | Decimal | Inches | Measure | 4-May |
| Weight | Weight of the arrestee | Decimal | Pounds | Measure | 145 |
| Hair | Hair color of the arrestee | Text/String | Alphanumeric | Dimension | Black |
| Eyes | Eye color of the arrestee | Text/String | Alphanumeric | Dimension | Brown |
| Race | Race of the arrestee | Text/String | Alphanumeric | Dimension | Hispanic or Latino |
| Ethnic | Ethnicity of the arrestee | Text/String | Alphanumeric | Dimension | Non-Hispanic or Latino |
| Sex | Gender of the arrestee | Text/String | Alphanumeric | Dimension | Male |
| Tatoo | Presence of tattoos on the arrestee | Text/String | Alphanumeric | Dimension | No |
| TatooComment | Comments or details about tattoos | Text/String | Alphanumeric | Dimension | NA |
| Occupation | Occupation of the arrestee | Text/String | Alphanumeric | Dimension | Unemployed |
| JobSchStatus | Job or school status of the arrestee | Text/String | Alphanumeric | Dimension | High School |
| Employer | Employer of the arrestee | Text/String | Alphanumeric | Dimension | NA |
| Drug | Presence of drugs on the arrestee | Text/String | Alphanumeric | Dimension | No |
| DrugRelated | Whether the arrest is drug-related | Text/String | Alphanumeric | Dimension | No |
| DrugType | Type of drug involved (if applicable) | Text/String | Alphanumeric | Dimension | NA |
| ClothingWorn | Description of clothing worn by the arrestee | Text/String | Alphanumeric | Dimension | YELLOW REFLECTIVE VEST,  BLUE JEANS, BOOTS |
| Expunged | Indicates whether the record has been expunged | Text/String | Alphanumeric | Dimension | No |
| UpZDate | Date of data update or revision | Date | YYYY-MM-DD | Dimension | 3/31/23 0:29 |

**Data Cleaning in Tableau Prep**

For the dataset I have chosen, I have performed data cleaning on Tableau Prep in the following ways.

Firstly, I have understood that city column has misspelled words for the names of cities. Hence, I have used Tableau Prep to correct them quickly and align them properly. I referred the names in the google and corrected with the best possible city name.

For example, As the cities are in and around Dallas, Grand Prairie has been written as Grand Pra, while Dallas named as DLS, Dalllas, etc.

A screenshot of a computer

Description automatically generated

Secondly, I have changed the role of few columns to their best possible. For example, the city column in the dataset has assigned the role with “City.” Likewise, I have changed role for State column to “State”.

A screenshot of a computer

Description automatically generated

In the third step, I can see that Zip code column has been assigned with integer. In general sense, we know zip code does not have any sensical meaning in statistical per se. Hence, I have changed the role of it to String.

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Finally, I have used filter to filter out city column containing nulls. Currently, I can think off about generating data visualizations based on city column. Hence there going to be issues if it contains nulls. Hence, I have filtered it out. However, I can perform further analysis and build strategy to handle null values in this context.

A screenshot of a computer

Description automatically generated

**Part 2**

Data visualization is the important and final step in the data science life cycle. How can one understand operations performed on data and outcome of it? Everyone likes to understand results quickly and completely. Hence data visualization step plays an outstanding role in the whole process of data analytics and science. It helps in showcasing patterns, insights and trends identified from the previous data processing steps and shows in the form of charts to make user understand and take decisions quickly.

Data visualization is very significant for any business. Stakeholder can get clear and crisp summary about complicated data and analytics. These reports can be foundational in solving business problems and take informed decisions. With the usage of interactive charts, a businessmen can deep dive into subtle and major points in the visualization which can help in brainstorming the cause and effects of the businesses. Visualization lets you to comprehend the huge amount of data in better way and helps in communicating results visually.

During my project in Data Analytics I course, I have worked on Layoffs analysis. The project about performing data analysis with layoffs made by various companies in diversified sectors. It is crucial to know the layoffs happened in different companies, sectors, and countries. Hence, I made multiple statistical and descriptive plots to make everyone understand the crux of the project with ease.

There are lots of benefits with data visualization, understanding trends and patterns, improved decision-making, real-time information about process and results, customized and interactive dashboards and visual perceptions about data.

Reference:

1. Article posted on splashBI.com; Article Name: Importance of Data Visualization; <https://splashbi.com/importance-purpose-benefit-of-data-visualization-tools/#:~:text=The%20importance%20of%20Data%20visualization,effective%20decisions%20quickly%20and%20accurately>.