

OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

C.LEANING DATASET

DASHBOARD

# Tableau Dataset Crime Data Los Angeles 2020-2024

Data Series 18.0 Business Intelligence Dibimbing

**Muhammad Aziz Habiburrahim**

HERE PORTOFOLIO MORE



[PORTOFOLIO WEBSITE](#)





# CONTENT



# CONTENT



OVERVIEW



IMPORT DATASET



EXPLORATORY DATASET



CLEANING DATASET



DASHBOARD

# OVERVIEW



## OVERVIEW

 ibimbing





# OVERVIEW

## DATASET

Dataset Crime Data from 2020 to Present adalah kumpulan data yang mencatat insiden kejahatan di Kota Los Angeles sejak tahun 2020 hingga saat ini. Data ini ditranskripsikan dari laporan kejahatan asli yang ditulis di atas kertas, sehingga mungkin terdapat beberapa ketidakakuratan dalam data tersebut. Beberapa kolom lokasi dengan data yang hilang dicatat sebagai (0°, 0°). Kolom alamat hanya disediakan hingga blok ratusan terdekat untuk menjaga privasi. Data ini seakurat data yang ada dalam basis data.

Dataset ini mencakup berbagai jenis kejahatan yang dilaporkan di Los Angeles, termasuk perampokan, penyerangan, pencurian, dan lainnya. Setiap entri dalam dataset biasanya mencakup informasi seperti tanggal dan waktu kejadian, jenis kejahatan, lokasi kejadian (dengan tingkat presisi hingga blok ratusan), dan informasi relevan lainnya. Dataset ini berguna untuk analisis tren kejahatan, identifikasi area dengan tingkat kejahatan tinggi, dan penelitian lainnya terkait keamanan publik.



# OVERVIEW

## JENIS KOLOM

KOLOM	TYPE	KOLOM	TYPE	KOLOM	TYPE
number_id	int64	crime_cd	int64	weapon_used_cd	float64
date_rptd	object	crime_cd_desc	object	weapon_desc	object
date_occ	object	mocodes	object	status	object
time_occ	int64	vict_age	int64	status_desc	object
area_number	int64	vict_sex	object	location	object
area_name	object	vict_descent	object	cross_street	object
rpt_dist_no	int64	premis_cd	float64	lat	float64
part_1/2	int64	premis_desc	object	lon	float64



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD



# IMPORT DATASET



IMPORT DATASET





# IMPORT DATASET

## IMPORT LIBRARY

Import ini menggunakan library pandas, matplotlib.pyplot, seaborn, drive, dan sheets pada platform [Google Colaboratory](#).

```
[24] # import library
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import drive
from google.colab import sheets
drive.mount('/content/drive')
```



Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).





# IMPORT DATASET

## LOAD DATASET

Load dataset yang sudah diimport, lalu tampilkan 5 baris pertama pada dataset.

```
[7] # read and load dataset
df = pd.read_csv('/content/drive/MyDrive/Course/Datasets/crime-data-los-angeles.csv')
df.head()
```



	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOCATION	Cross Street	LAT	LON
0	190326475	03/01/2020 12:00:00 AM	03/01/2020 12:00:00 AM	2130	7	Wilshire	784	1	510	VEHICLE - STOLEN	...	AA	Adult Arrest	510.0	998.0	NaN	NaN	1900 S LONGWOOD AV	NaN	34.0375	-118.3506
1	200106753	02/09/2020 12:00:00 AM	02/08/2020 12:00:00 AM	1800	1	Central	182	1	330	BURGLARY FROM VEHICLE	...	IC	Invest Cont	330.0	998.0	NaN	NaN	1000 S FLOWER ST	NaN	34.0444	-118.2628
2	200320258	11/11/2020 12:00:00 AM	11/04/2020 12:00:00 AM	1700	3	Southwest	356	1	480	BIKE - STOLEN	...	IC	Invest Cont	480.0	NaN	NaN	NaN	1400 W 37TH ST	NaN	34.0210	-118.3002
3	200907217	05/10/2023 12:00:00 AM	03/10/2020 12:00:00 AM	2037	9	Van Nuys	964	1	343	SHOPLIFTING-GRAND THEFT (\$950.01 & OVER)	...	IC	Invest Cont	343.0	NaN	NaN	NaN	14000 RIVERSIDE DR	NaN	34.1576	-118.4387
4	200412582	09/09/2020 12:00:00 AM	09/09/2020 12:00:00 AM	630	4	Hollenbeck	413	1	510	VEHICLE - STOLEN	...	IC	Invest Cont	510.0	NaN	NaN	NaN	200 E AVENUE 28	NaN	34.0820	-118.2130

5 rows × 28 columns

OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD





# IMPORT DATASET

## LOAD DATASET

Tampilkan juga baris 5 baris terakhir untuk mengetahui berapa banyak baris pada dataset. Disini terlihat ada sebanyak 1.004.846 baris.

```
[8] # display 5 last rows  
df.tail()
```



	DR_NO	Date Rptd	DATE OCC	TIME OCC	AREA	AREA NAME	Rpt Dist No	Part 1-2	Crm Cd	Crm Cd Desc	...	Status	Status Desc	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOCATION	Cross Street	LA
1004842	240710284	07/24/2024 12:00:00 AM	07/23/2024 12:00:00 AM	1400	7	Wilshire	788	1	510	VEHICLE - STOLEN	...	IC	Invest Cont	510.0	NaN	NaN	NaN	4000 W 23RD ST	NaN	34.036
1004843	240104953	01/15/2024 12:00:00 AM	01/15/2024 12:00:00 AM	100	1	Central	101	2	745	VANDALISM - MISDEAMEANOR (\$399 OR UNDER)	...	IC	Invest Cont	745.0	NaN	NaN	NaN	1300 W SUNSET BL	NaN	34.068
1004844	240410786	10/14/2024 12:00:00 AM	10/11/2024 12:00:00 AM	2330	4	Hollenbeck	421	1	341	THEFT-GRAND (\$950.01 & OVER)EXCPT,GUNS,FOWL,LI...	...	IC	Invest Cont	341.0	NaN	NaN	NaN	1700 ALBION ST	NaN	34.067
1004845	240309674	04/24/2024 12:00:00 AM	04/24/2024 12:00:00 AM	1500	3	Southwest	358	1	230	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	...	IC	Invest Cont	230.0	NaN	NaN	NaN	FLOWER ST	JEFFERSON BL	34.021
1004846	240910892	08/13/2024 12:00:00 AM	08/12/2024 12:00:00 AM	2300	9	Van Nuys	914	1	510	VEHICLE - STOLEN	...	IC	Invest Cont	510.0	NaN	NaN	NaN	6900 VESPER AV	NaN	34.196

5 rows × 28 columns

OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD



# EXPLORATORY DATASET



EXPLORATORY DATASET





# EXPLORATORY DATASET

```
[9] # display info type data columns  
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1004847 entries, 0 to 1004846  
Data columns (total 28 columns):  
#   Column                Non-Null Count  Dtype    
---  ---  
0   DR_NO                 1004847 non-null int64    
1   Date Rptd            1004847 non-null object   
2   DATE OCC              1004847 non-null object   
3   TIME OCC              1004847 non-null int64    
4   AREA                  1004847 non-null int64    
5   AREA NAME             1004847 non-null object   
6   Rpt Dist No           1004847 non-null int64    
7   Part 1-2              1004847 non-null int64    
8   Crm Cd                1004847 non-null int64    
9   Crm Cd Desc           1004847 non-null object   
10  Mocodes               853256 non-null object   
11  Vict Age              1004847 non-null int64    
12  Vict Sex              860243 non-null object   
13  Vict Descent          860231 non-null object   
14  Premis Cd             1004831 non-null float64   
15  Premis Desc           1004259 non-null object   
16  Weapon Used Cd        327193 non-null float64   
17  Weapon Desc           327193 non-null object   
18  Status                1004846 non-null object   
19  Status Desc           1004847 non-null object   
20  Crm Cd 1              1004836 non-null float64   
21  Crm Cd 2              69130 non-null float64   
22  Crm Cd 3              2315 non-null float64   
23  Crm Cd 4              64 non-null float64   
24  LOCATION              1004847 non-null object   
25  Cross Street          154221 non-null object   
26  LAT                   1004847 non-null float64   
27  LON                   1004847 non-null float64   
dtypes: float64(8), int64(7), object(13)  
memory usage: 214.7+ MB
```

## INFO DATASET

Berikut terdapat informasi mengenai data seperti variabel kolom, jumlah data tidak kosong, dan tipe datanya.





# EXPLORATORY DATASET

## INFO DATA STATISTIK

Operasikan statistik pada dataset untuk melihat count, mean, standar deviasi, minimal, 25%, 50%, 75% dan maximum.

```
[10] # summary statistics dataset  
df.describe()
```



	DR_NO	TIME OCC	AREA	Rpt Dist No	Part 1-2	Crm Cd	Vict Age	Premis Cd	Weapon Used Cd	Crm Cd 1	Crm Cd 2
count	1.004847e+06	1.004847e+06	1.004847e+06	1.004847e+06	1.004847e+06	1.004847e+06	1.004847e+06	1.004831e+06	327193.000000	1.004836e+06	69130.000000
mean	2.202171e+08	1.339911e+03	1.069147e+01	1.115607e+03	1.400307e+00	5.001353e+02	2.891905e+01	3.056073e+02	363.944543	4.998958e+02	958.115941
std	1.318835e+07	6.510691e+02	6.110286e+00	6.111622e+02	4.899607e-01	2.052622e+02	2.199224e+01	2.192915e+02	123.741225	2.050627e+02	110.359911
min	8.170000e+02	1.000000e+00	1.000000e+00	1.010000e+02	1.000000e+00	1.100000e+02	-4.000000e+00	1.010000e+02	101.000000	1.100000e+02	210.000000
25%	2.106168e+08	9.000000e+02	5.000000e+00	5.870000e+02	1.000000e+00	3.310000e+02	0.000000e+00	1.010000e+02	311.000000	3.310000e+02	998.000000
50%	2.209158e+08	1.420000e+03	1.100000e+01	1.139000e+03	1.000000e+00	4.420000e+02	3.000000e+01	2.030000e+02	400.000000	4.420000e+02	998.000000
75%	2.311101e+08	1.900000e+03	1.600000e+01	1.613000e+03	2.000000e+00	6.260000e+02	4.400000e+01	5.010000e+02	400.000000	6.260000e+02	998.000000
max	2.521040e+08	2.359000e+03	2.100000e+01	2.199000e+03	2.000000e+00	9.560000e+02	1.200000e+02	9.760000e+02	516.000000	9.560000e+02	999.000000

OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD



# EXPLORATORY DATASET

```
# checking for missing values
df.isnull().sum()
```

	0
DR_NO	0
Date Rptd	0
DATE OCC	0
TIME OCC	0
AREA	0
AREA NAME	0
Rpt Dist No	0
Part 1-2	0
Crm Cd	0
Crm Cd Desc	0
Mocodes	151591
Vict Age	0
Vict Sex	144604
Vict Descent	144616
Premis Cd	16
Premis Desc	588
Weapon Used Cd	677654
Weapon Desc	677654
Status	1
Status Desc	0
Crm Cd 1	11
Crm Cd 2	935717
Crm Cd 3	1002532
Crm Cd 4	1004783
LOCATION	0
Cross Street	850626
LAT	0
LON	0

## INFO DATA KOSONG

Berikut menampilkan data kosong pada dataset, disini terlihat ada kolom yang terdapat data kosong seperti Mocodes, Vict Sex, Vict Descent, Premis Cd, Premis Desc, Weapon Used Cd, Weapon Desc, Status, Crime Cd 1-4, dan Cross Street.





# EXPLORATORY DATASET

## INFO DATA DUPLIKAT

Berikut menampilkan data yang duplikat pada dataset. Disini terlihat tidak ada satupun data yang duplikat dengan jumlah 0 data.

```
[12] # checking for duplicates
      duplicates = df.duplicated().sum()
      print("\nNumber of duplicates in the dataset:")
      print(duplicates)
```



```
Number of duplicates in the dataset:
0
```



# CLEANING DATASET



CLEANING DATASET

 ibimbing





# CLEANING DATASET

```
[13] # rename variable
df = df.rename(columns={'DR_NO': 'number_id',
                        'Date Rptd': 'date_rptd',
                        'DATE OCC' : 'date_occ',
                        'TIME OCC' : 'time_occ',
                        'AREA': 'area_number',
                        'AREA NAME': 'area_name',
                        'Rpt Dist No' : 'rpt_dist_no',
                        'Part 1-2' : 'part_1/2',
                        'Crm Cd' : 'crime_cd',
                        'Crm Cd Desc': 'crime_cd_desc',
                        'Mocodes' : 'mocodes',
                        'Vict Age' : 'vict_age',
                        'Vict Sex': 'vict_sex',
                        'Vict Descent': 'vict_descent',
                        'Premis Cd' : 'premis_cd',
                        'Premis Desc' : 'premis_desc',
                        'Weapon Used Cd': 'weapon_used_cd',
                        'Weapon Desc': 'weapon_desc',
                        'Status' : 'status',
                        'Status Desc' : 'status_desc',
                        'Crm Cd 1' : 'crime_cd_1',
                        'Crm Cd 2' : 'crime_cd_2',
                        'Crm Cd 3' : 'crime_cd_3',
                        'Crm Cd 4' : 'crime_cd_4',
                        'LOCATION' : 'location',
                        'Cross Street': 'cross_street',
                        'LAT' : 'lat',
                        'LON' : 'lon'})

df.head()
```

## RENAME VARIABEL KOLOM

Mengubah nama variabel pada kolom dataset dengan catatan tidak ada huruf besar dan spasi diganti oleh garis bawah.







# CLEANING DATASET

## REPLACE - REMOVE VARIABEL DATASET

Mengubah kolom data kosong menjadi data tak kosong dengan menggunakan mode pada object dan median pada numerik dan menghapus kolom crime cd 1-4.

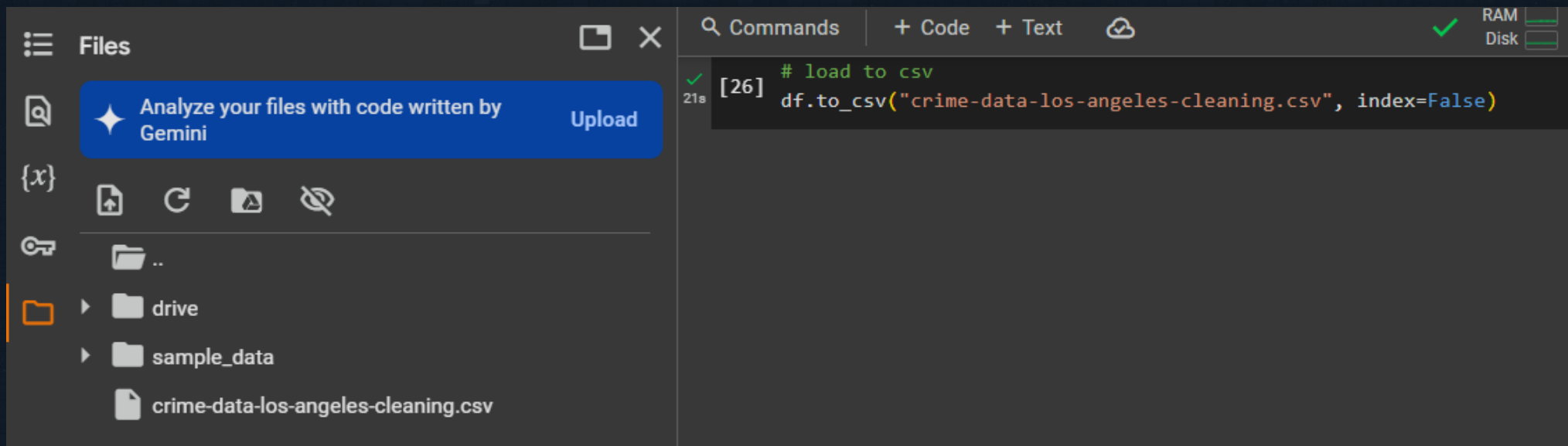
```
[15] # replace and columns missing values
df['mocodes'].fillna(df['mocodes'].mode()[0], inplace=True)
df['vict_sex'].fillna(df['vict_sex'].mode()[0], inplace=True)
df['vict_descent'].fillna(df['vict_descent'].mode()[0], inplace=True)
df['premis_cd'].fillna(df['premis_cd'].median(), inplace=True)
df['premis_desc'].fillna(df['premis_desc'].mode()[0], inplace=True)
df['weapon_used_cd'].fillna(df['weapon_used_cd'].median(), inplace=True)
df['weapon_desc'].fillna(df['weapon_desc'].mode()[0], inplace=True)
df['status'].fillna(df['status'].mode()[0], inplace=True)
df['cross_street'].fillna(df['cross_street'].mode()[0], inplace=True)
df.drop(columns=['crime_cd_1'], inplace=True)
df.drop(columns=['crime_cd_2'], inplace=True)
df.drop(columns=['crime_cd_3'], inplace=True)
df.drop(columns=['crime_cd_4'], inplace=True)
```



# CLEANING DATASET

## LOAD TO CSV CLEANING

Setelah cleaning pada dataset, lalu load kembali dalam bentuk csv dengan nama file crime-data-los-angeles-cleaning.csv dan download pada files pada gambar berikut untuk di olah menggunakan dashboard.





# DASHBOARD



## DASHBOARD

 ibimbing



# DASHBOARD

## IMPORT DATASET

Dashboard ini menggunakan software [Tableau](#) yang memudahkan untuk memvisualisasikan data. Berikut tampilan dataset pada halaman utama tableau dengan import dataset terlebih dahulu.

Name			
crime-data-los-angeles-cleaning.csv			
Fields			
Type	Field Name	Physical Table	Remote Field Name
#	Number Id	crime-data-los-angeles-cleaning.csv	number_id
📅	Date Rptd	crime-data-los-angeles-cleaning.csv	date_rptd
📅	Date Occ	crime-data-los-angeles-cleaning.csv	date_occ
#	Time Occ	crime-data-los-angeles-cleaning.csv	time_occ
#	Area Number	crime-data-los-angeles-cleaning.csv	area_number
Abc	Area Name	crime-data-los-angeles-cleaning.csv	area_name
#	Rpt Dist No	crime-data-los-angeles-cleaning.csv	rpt_dist_no
#	Part 1/2	crime-data-los-angeles-cleaning.csv	part_1/2
#	Crime Cd	crime-data-los-angeles-cleaning.csv	crime_cd
Abc	Crime Cd Desc	crime-data-los-angeles-cleaning.csv	crime_cd_desc

OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

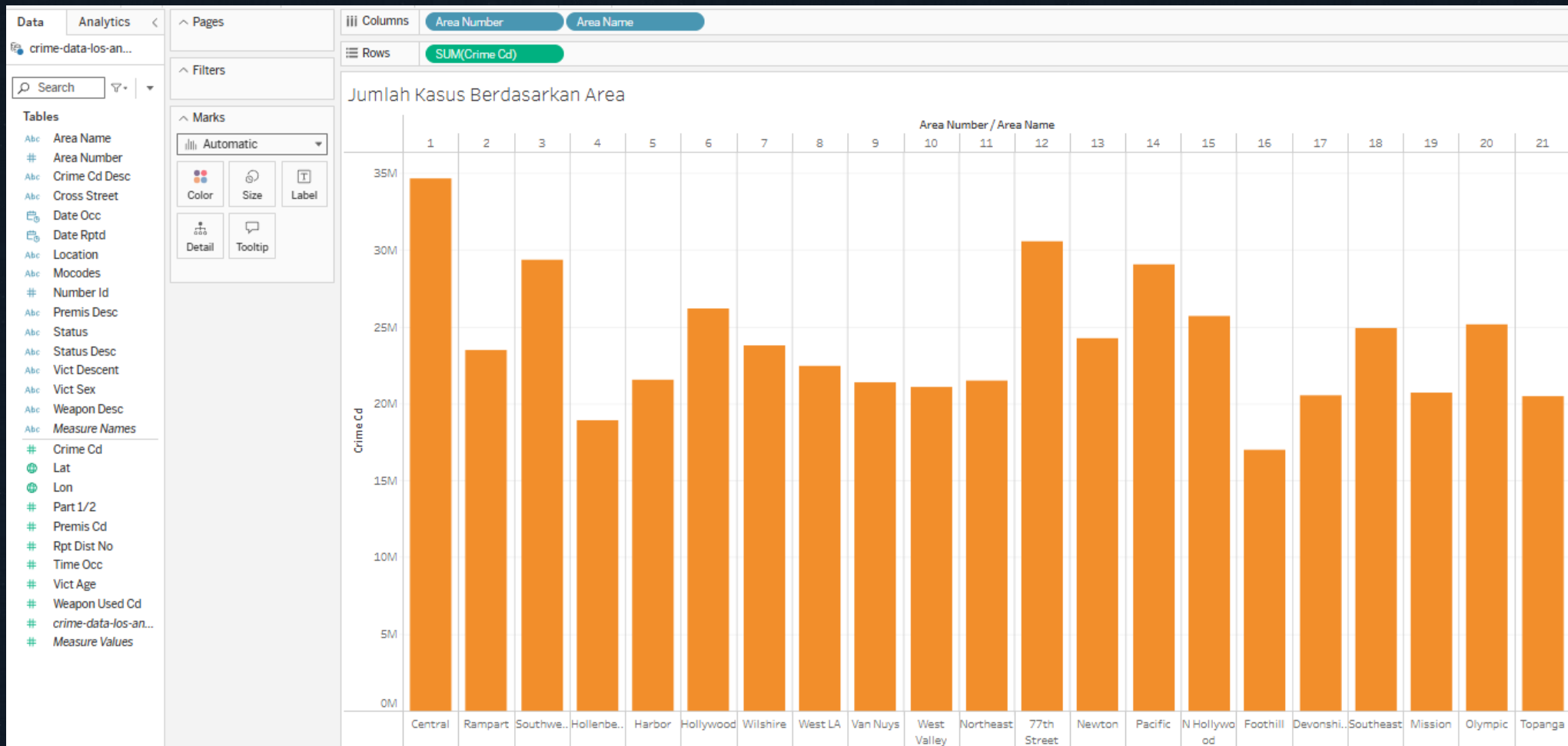
DASHBOARD





# DASHBOARD

AREA NUMBER/AREA NAME – CRIME CD  
Central merupakan area yang paling banyak kasus kriminal.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

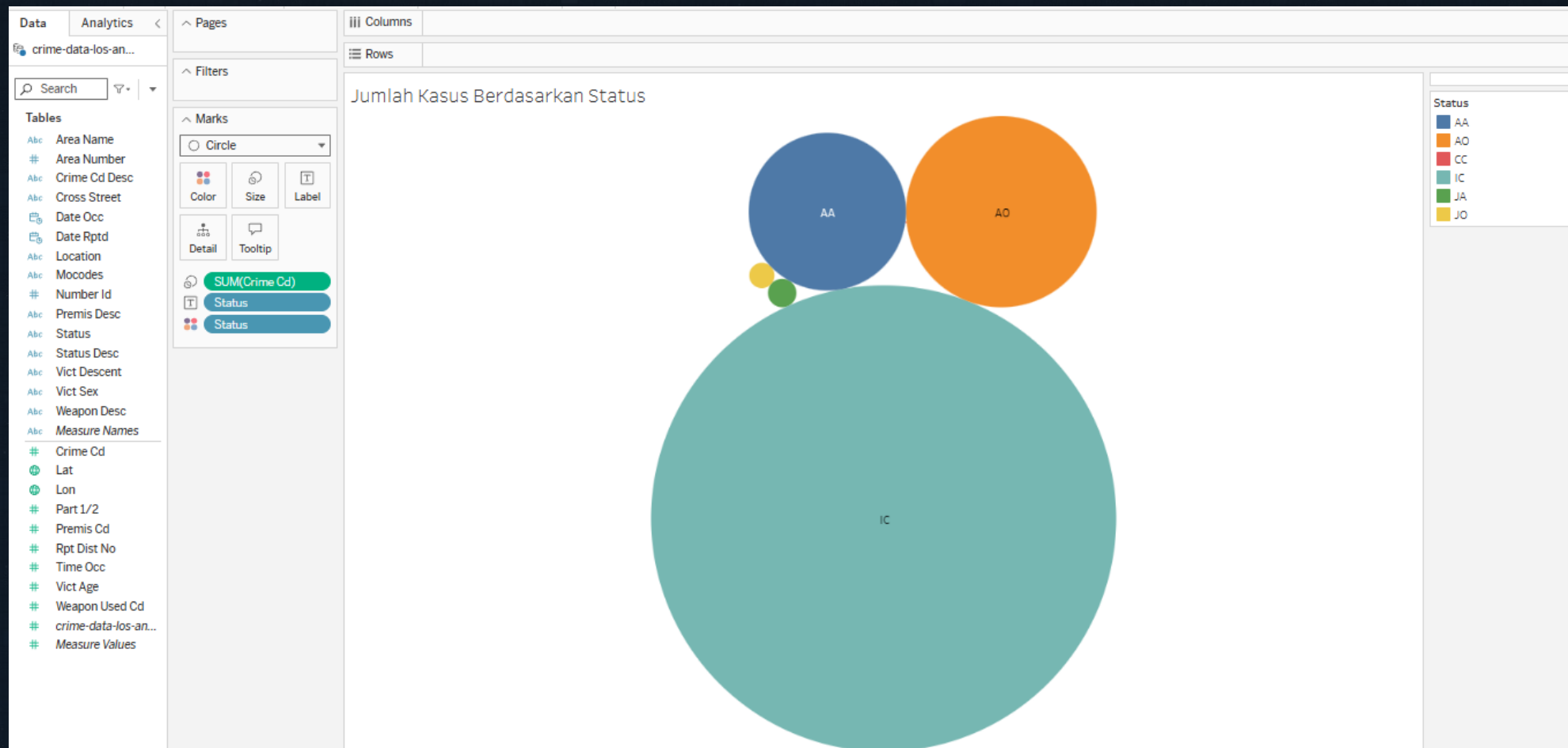
DASHBOARD



# DASHBOARD

## STATUS – CRIME CD

IC merupakan status yang paling banyak kasus kriminal.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD

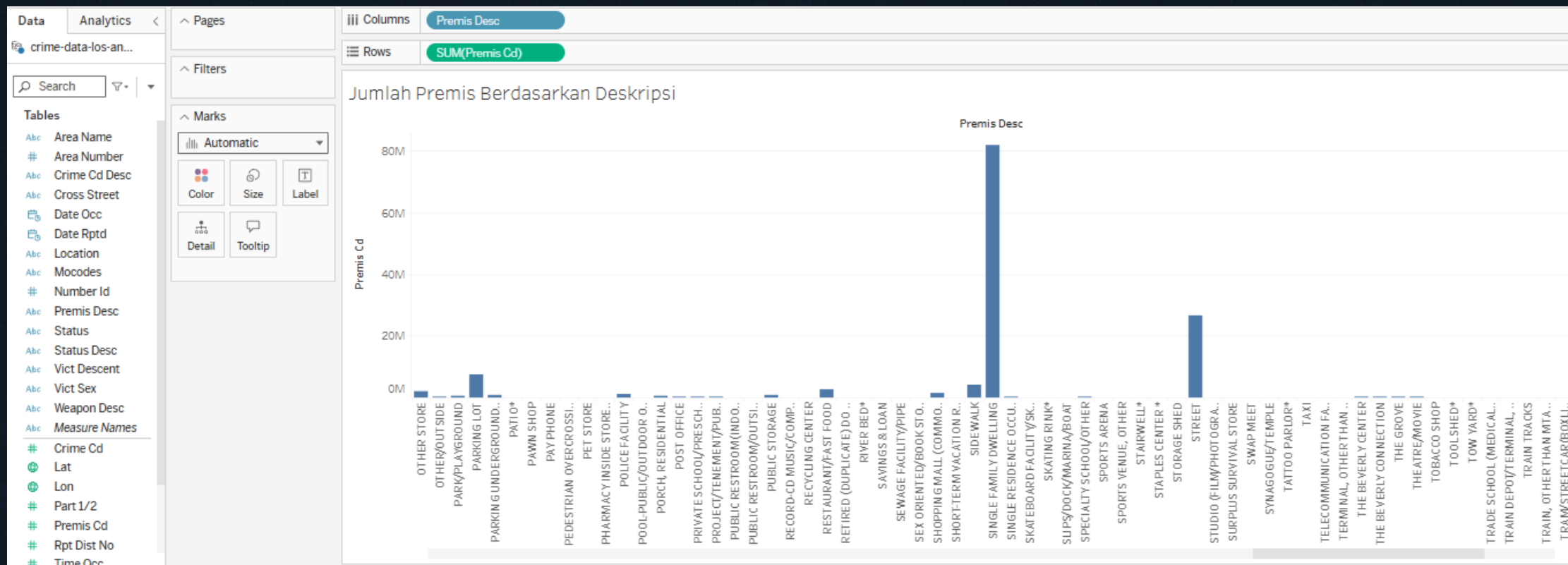




# DASHBOARD

## PREMIS DESC – CRIME CD

Single Family Dwelling merupakan premis yang paling banyak kasus kriminal.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

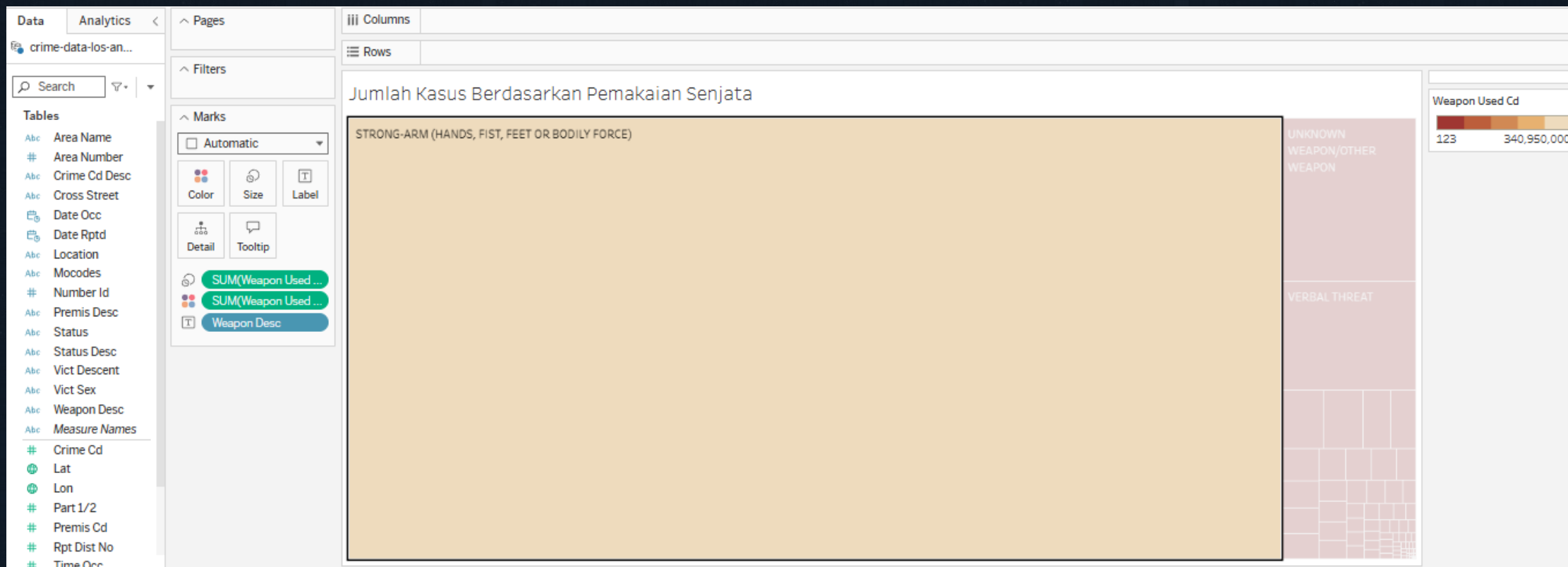
DASHBOARD



# DASHBOARD

## WEAPON DESC – WEAPON USED CD

STRONG ARM merupakan senjata yang paling banyak kasus kriminal.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD

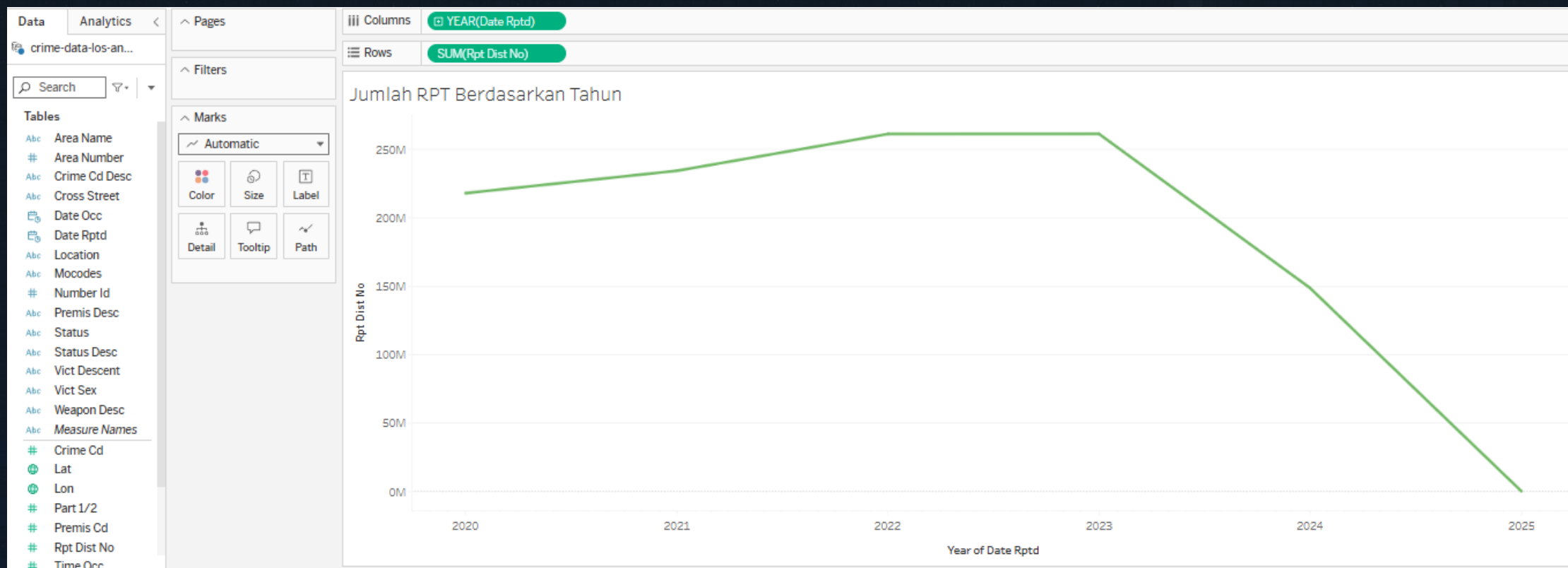




# DASHBOARD

## DATE RPTD – NO RPTD

Sebelum 2023 RPTD mengalami kenaikan dan setelah 2023 RPTD mengalami penurunan.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

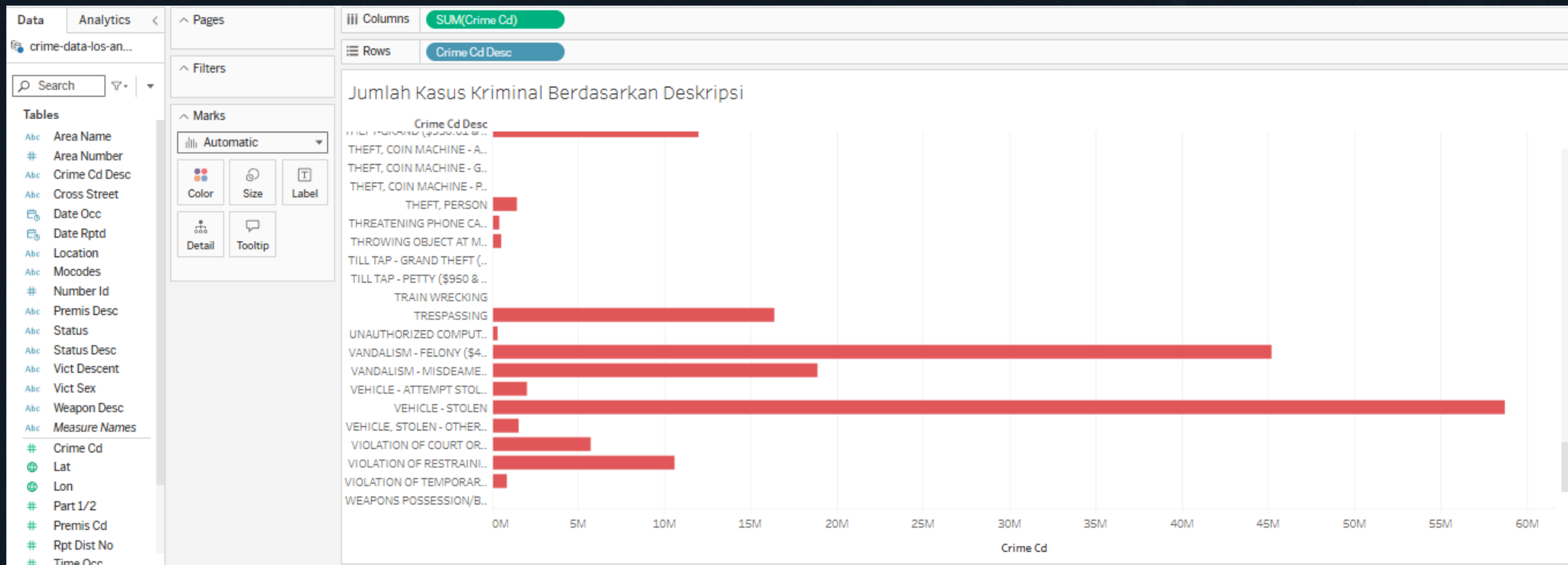
DASHBOARD



# DASHBOARD

## CRIME CD DESC – CRIME CD

VEHICLE - STOLEN merupakan jenis kasus yang paling banyak.



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD

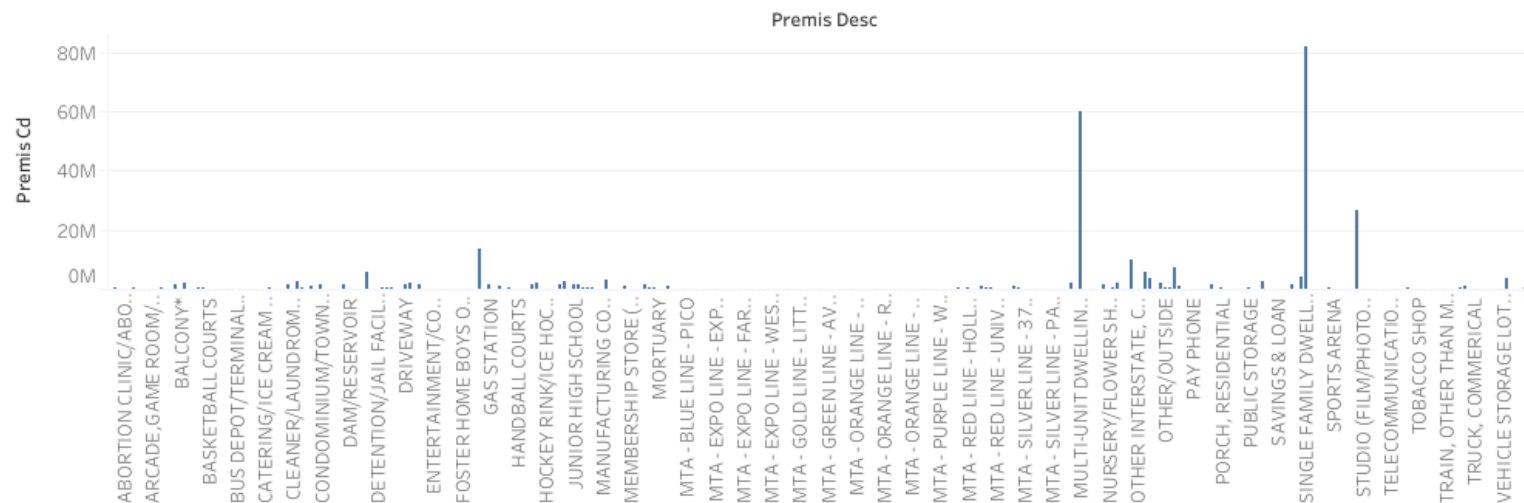




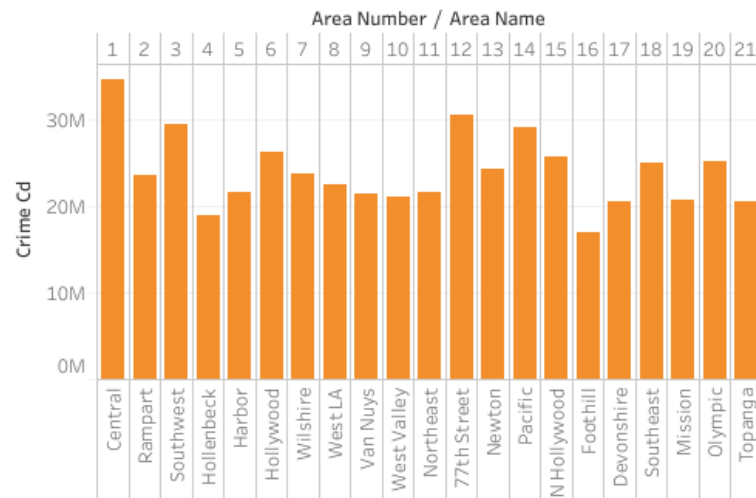
# DASHBOARD

## DASHBOARD 1

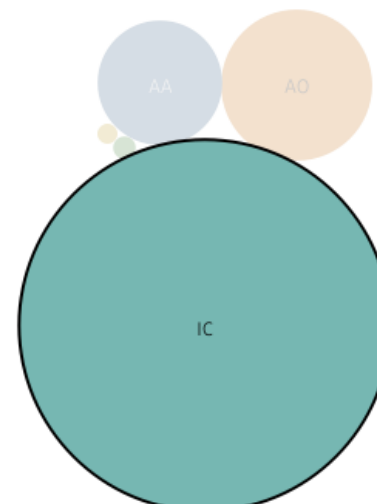
Jumlah Premis Berdasarkan Deskripsi



Jumlah Kasus Berdasarkan Area



Jumlah Kasus Berdasarkan Status

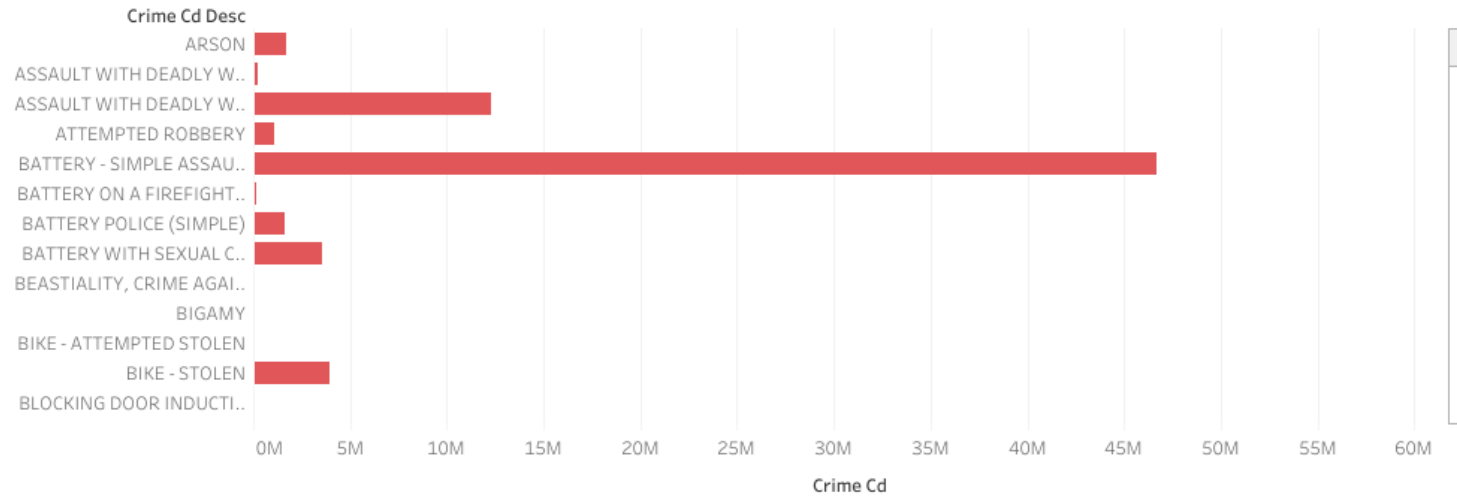




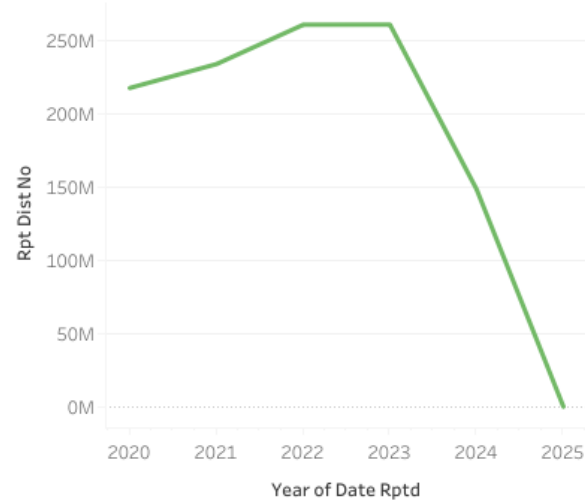
# DASHBOARD

## DASHBOARD 2

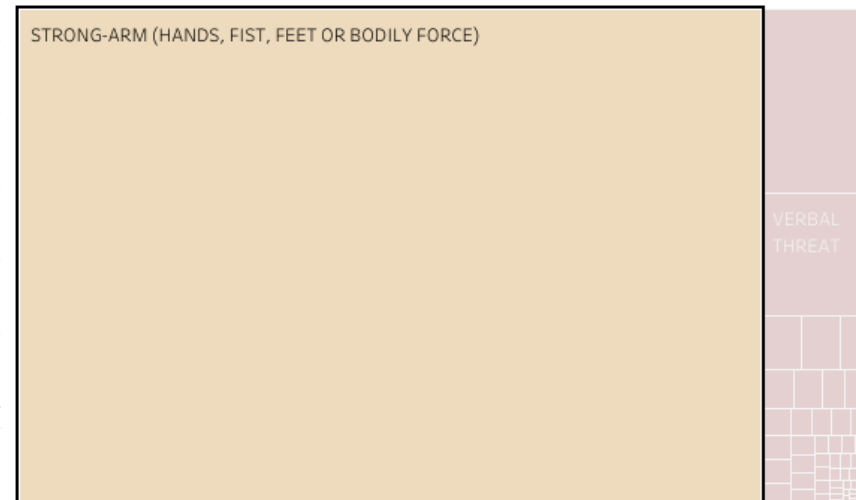
Jumlah Kasus Kriminal Berdasarkan Deskripsi



Jumlah RPT Berdasarkan Tahun



Jumlah Kasus Berdasarkan Pemakaian Senjata



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD



OVERVIEW

IMPORT DATASET

EXPLORATORY DATASET

CLEANING DATASET

DASHBOARD



# THANKS FOR LISENING!



SEE YOU NEXT TIME!

## Muhammad Aziz Habiburrahim

HERE PORTOFOLIO MORE



[PORTOFOLIO WEBSITE](#)