

CLEANING DATA IN POWER QUERY

Crime Data Analyst Svetlana Gubin,
Oakland PD

International Association Crime Analysts
March 15, 2023

Instructor & Course Introduction

Svetlana Gubin, Oakland PD (510)238-2017

Sgubin@OaklandCA.gov



Power Query

- ✓ Power Query documentation https://docs.microsoft.com/en-us/power-query
- ✓ Power Query M Language documentation https://learn.microsoft.com/en-us/powerquery-m-language-specification



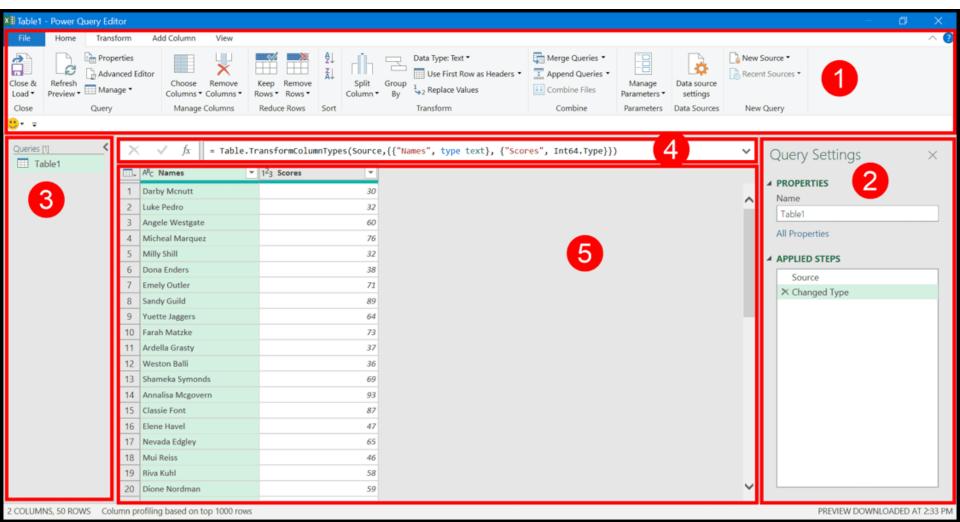
CONTENTS

- ✓ Cleaning Data in Power Query
 - Date/time cleaning
 - Text cleaning
 - Number cleaning

✓ Additional Information

✓ Conclusion and QA

Power Query – Power Query Editor



1. Ribbon

2. Applied Steps

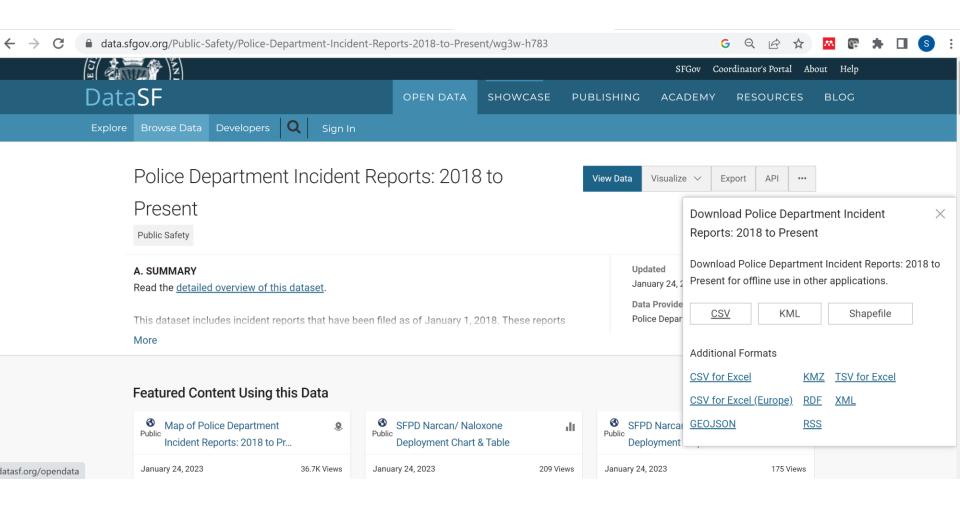
3. Queries

4. Formula Bar

5. Data Preview

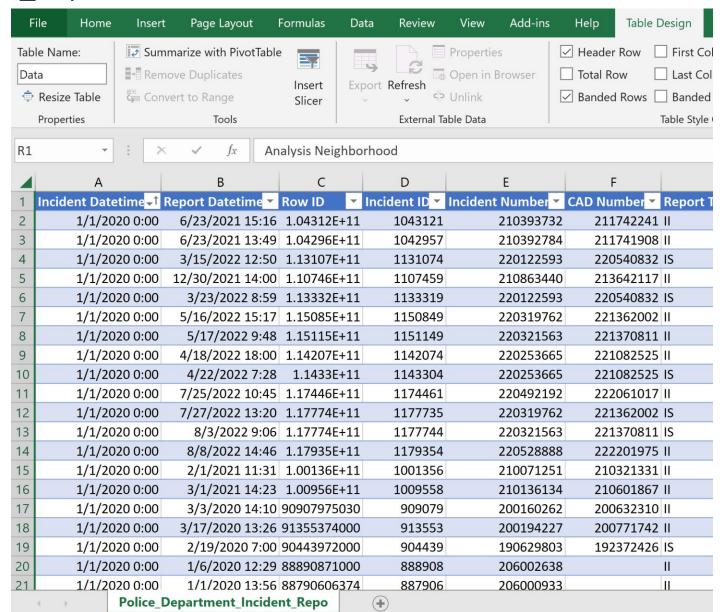
https://excelchamps.com/power-query/

DATABASE SAN FRANCISCO

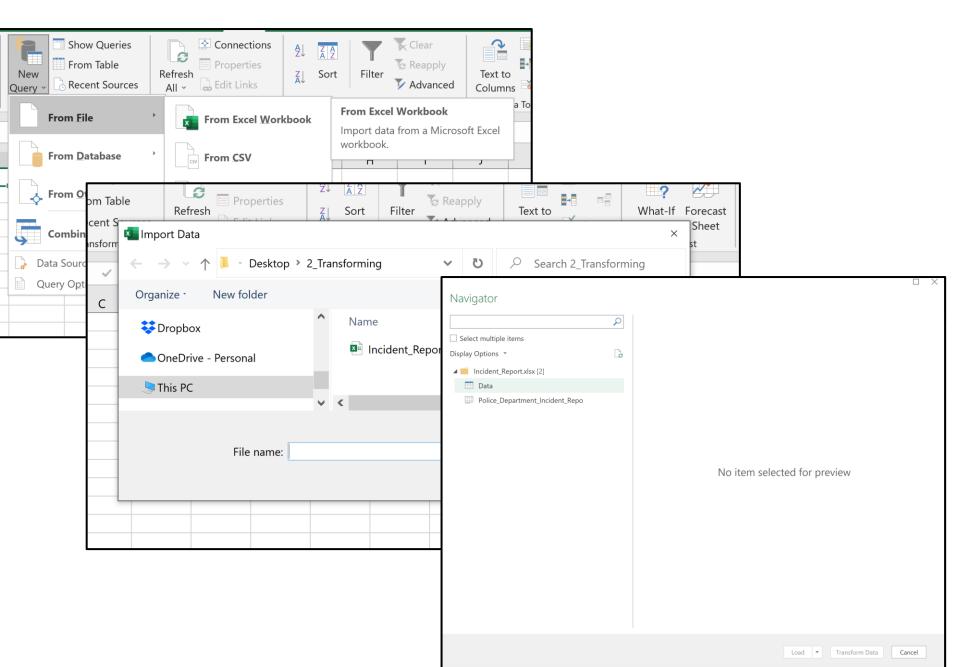


DATA

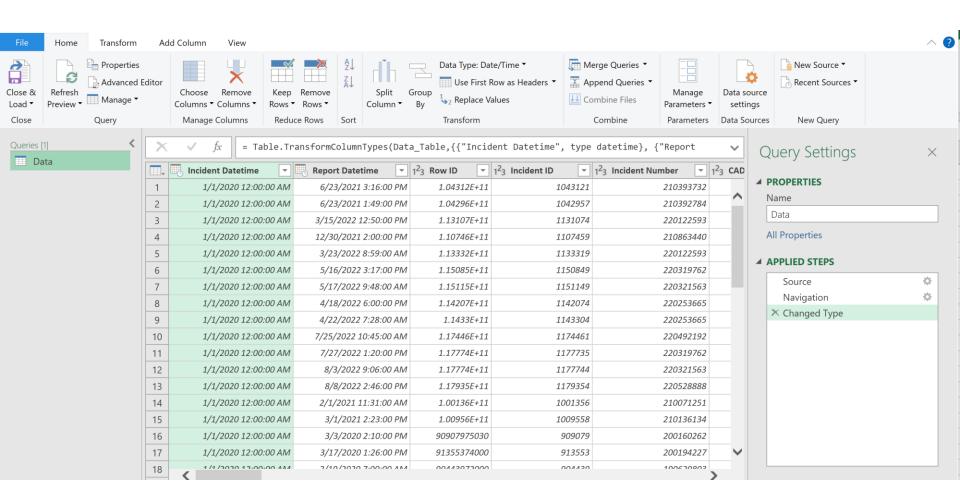
Incident_Report.xlsx



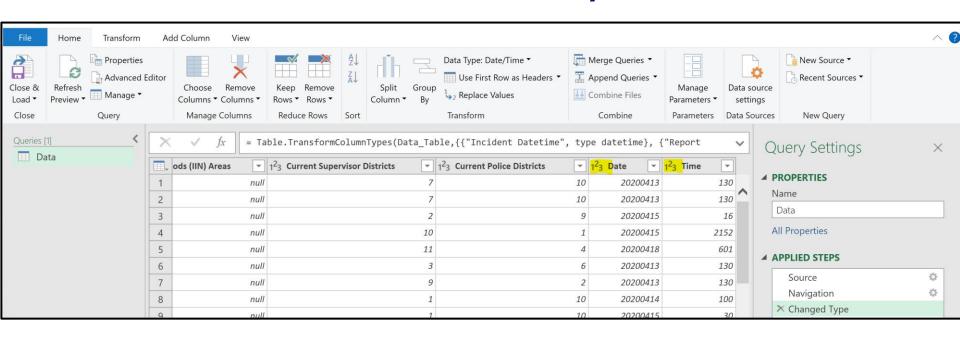
ADD DATA

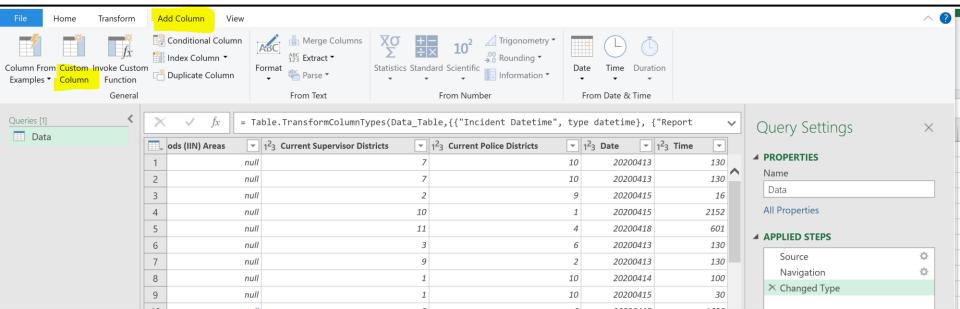


TRANSFORM DATA - POWER QUERY EDITOR



DATE CLEANING -DATE AS NUMBERS/TEXT





DATE CLEANING - DATE AS NUMBERS

Custom Column

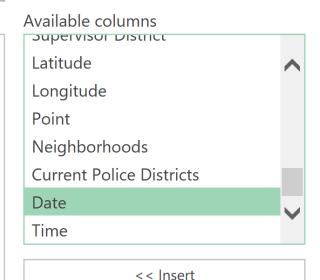
Add a column that is computed from the other columns.

New column name

DateFormatted

Custom column formula

= Date.FromText(Number.ToText([Date]))



Learn about Power Query formulas

✓ No syntax errors have been detected.

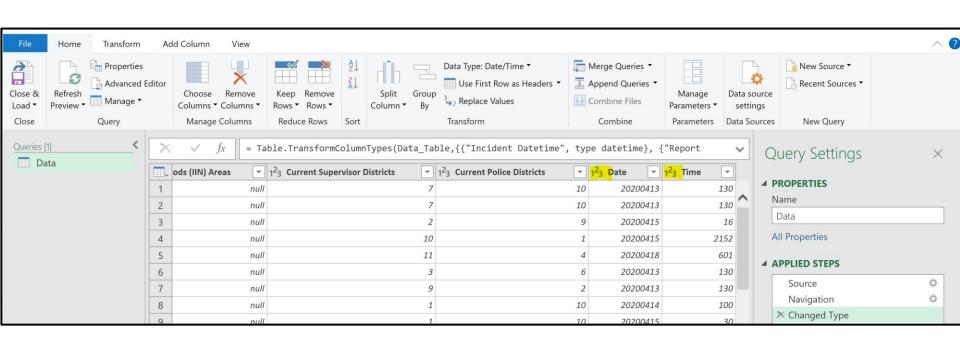
OK

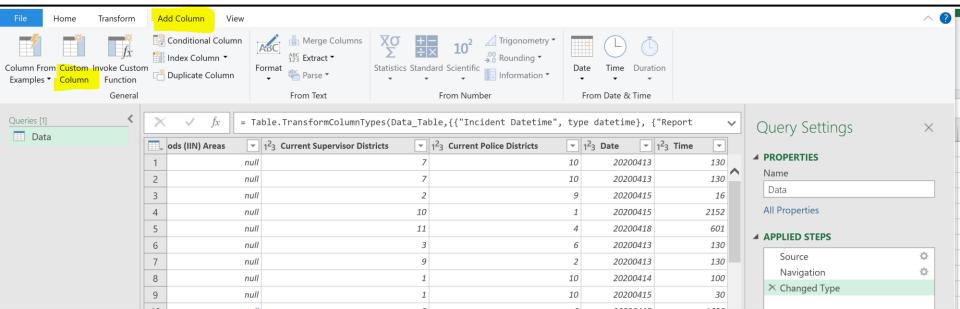
Cancel

DATE CLEANING - DATE AS NUMBERS

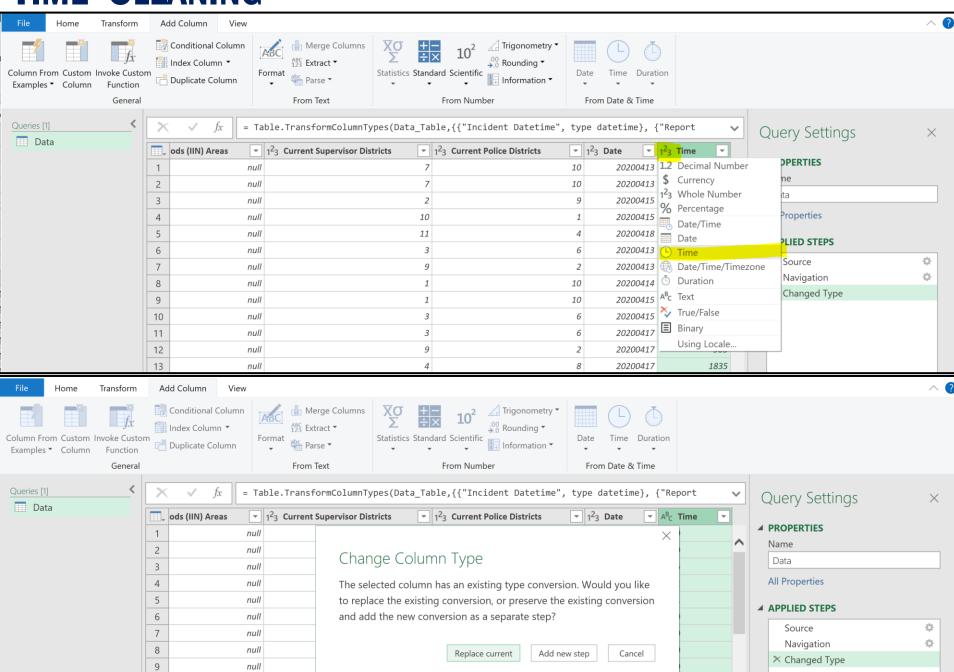
Hom Number Hom Date & Time					
"DateFormatted", each Date.FromText(Number.ToText					
cts	1 ² ₃ Date ▼	1 ² ₃ Time ▼	ABC DateFormatted		
10	20200413	130	4/13/2020		
				Name	
10	20200413	130	4/13/2020	Data	
				All Pr	
9	20200415	16	4/15/2020	⊿ APPL	
1	20200415	2152	4/15/2020		
4	20200418	601	4/18/2020	S	
6	20200413	130	4/13/2020	l N	
2	20200413	130	4/13/2020	× A	
10	20200414	100	4/14/2020	C	
10	20200415	30	4/15/2020		
6	20200415	1600	4/15/2020		
6	20200417	825	4/17/2020		
2	20200417	903	4/17/2020		
8	20200417	1835	4/17/2020		
3	null	null	null		
7	null	null	nullnull	~	
10	lling	lling	null		

TIME CLEANING

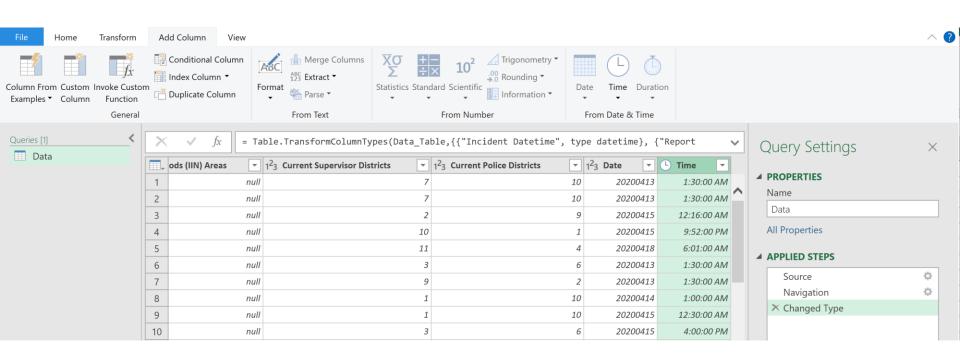




TIME CLEANING



TIME CLEANING



CLEANING -NULL

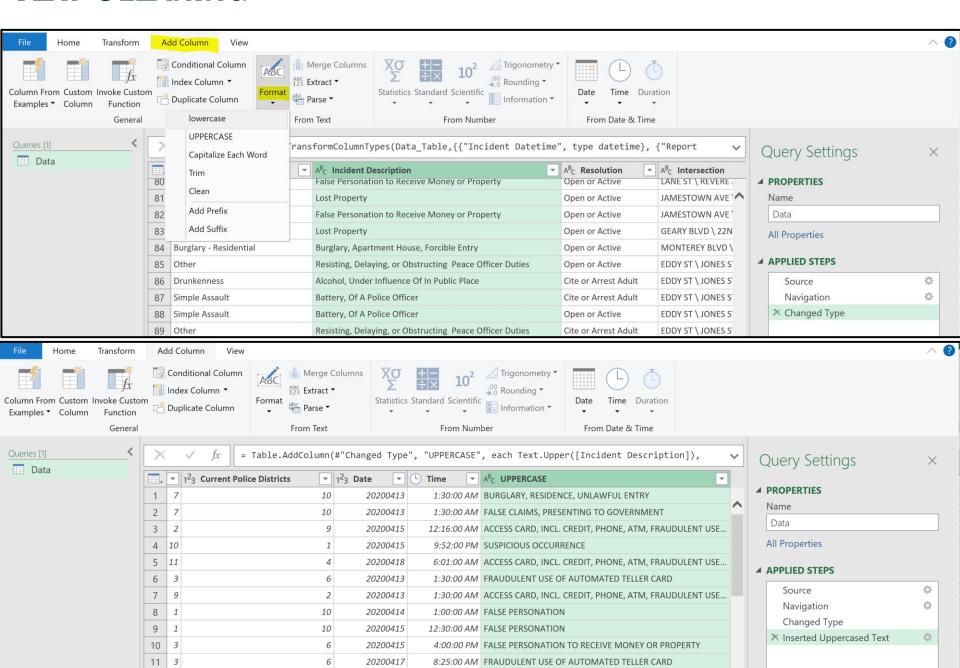
Nulls is a special symbol that means either unknown or inapplicable.

It does not mean zero or blank!!!!

Can create problems when:

- functions such as COUNT, AVERAGE, and SUM are used
- with date/ time formatting

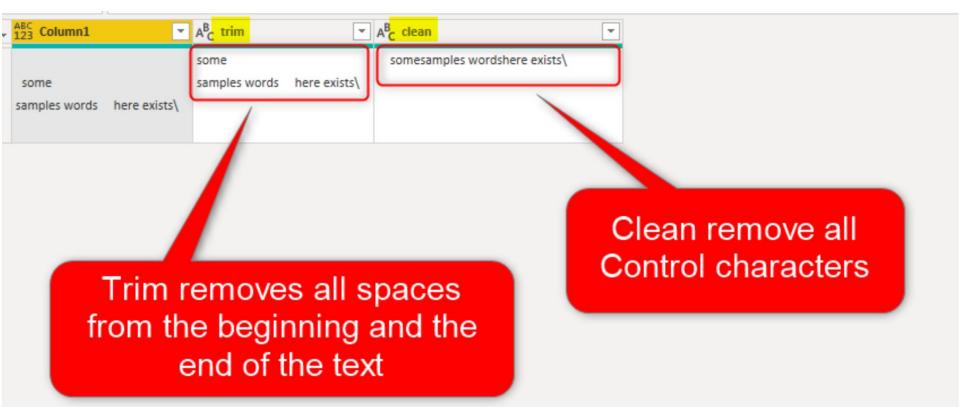
TEXT CLEANING



TEXT CLEANING - TRIM VS CLEAN

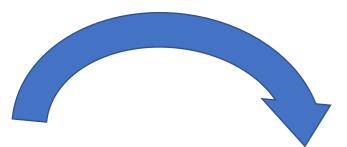
Trim removes all the spaces from the beginning and end of the text values.

Clean removes all the control characters from the text value. This can include end of lines, tabs etc.



https://radacad.com/trim-vs-clean-in-power-bi-and-power-query

CLEANING TEXT - TRIM



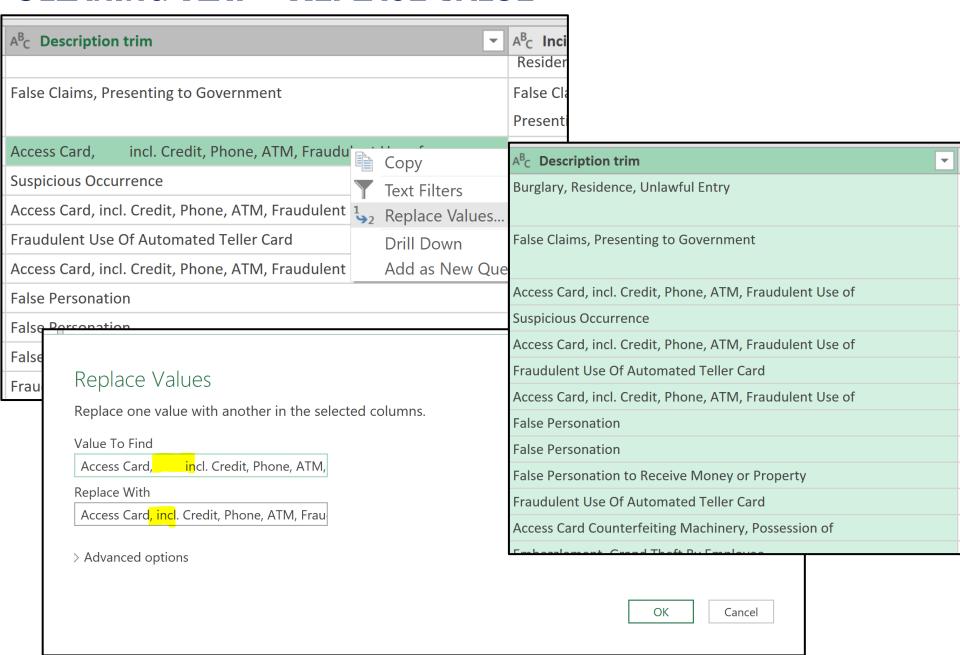
		<u> </u>
A ^B _C Description trim	A	A ^B _C Description trim
Burglary,	В	Burglary,
Residence, Unlawful Entry	F	Residence, Unlawful Entry
False Claims,	Fi	False Claims,
Presenting to Government	Р	Presenting to Government
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Α	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of
Suspicious Occurrence	Sı	Suspicious Occurrence
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Α	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of
Fraudulent Use Of Automated Teller Card	Fi	Fraudulent Use Of Automated Teller Card
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Α	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of
False Personation	Fi	False Personation
False Personation	Fi	False Personation
False Personation to Receive Money or Property	Fi	False Personation to Receive Money or Property
Fraudulent Use Of Automated Teller Card	Fi	Fraudulent Use Of Automated Teller Card
Access Card Counterfeiting Machinery, Possession of	Α	Access Card Counterfeiting Machinery, Possession of
Embezzlement, Grand Theft By Employee	E	Embezzlement, Grand Theft By Employee
Miscellaneous Investigation	N	Miscellaneous Investigation
Suspicious Occurrence	Sı	Suspicious Occurrence
0 0		

CLEANING TEXT - CLEANING

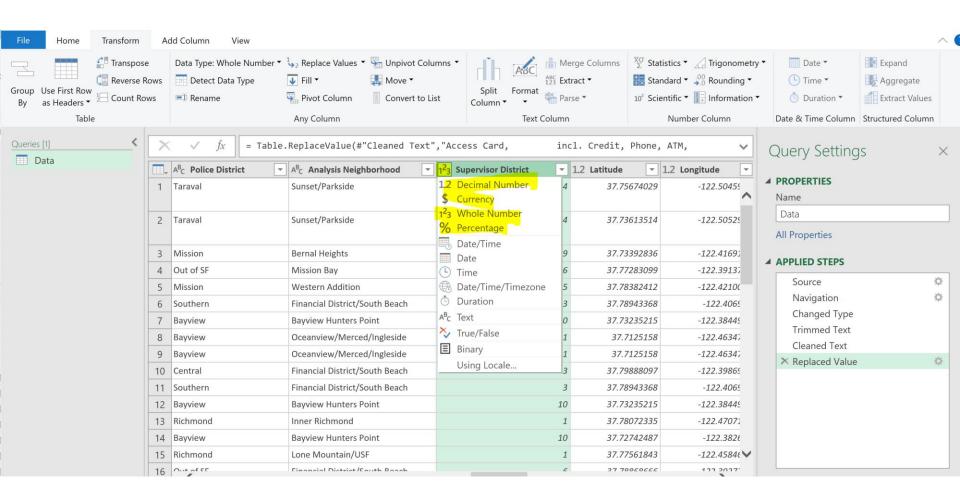


A ^B _C Description trim	A ^B C Description trim			
Burglary,	Burglary, Residence, Unlawful Entry			
Residence, Unlawful Entry				
False Claims,	False Claims, Presenting to Government			
Presenting to Government				
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of			
Suspicious Occurrence	Suspicious Occurrence			
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of			
Fraudulent Use Of Automated Teller Card	Fraudulent Use Of Automated Teller Card			
Access Card, incl. Credit, Phone, ATM, Fraudulent Use of	Access Card, incl. Credit, Phone, ATM, Fraudulent Use of			
False Personation	False Personation			
False Personation	False Personation			
False Personation to Receive Money or Property	False Personation to Receive Money or Property			
Fraudulent Use Of Automated Teller Card	Fraudulent Use Of Automated Teller Card			
Access Card Counterfeiting Machinery, Possession of	Access Card Counterfeiting Machinery, Possession of			
Embezzlement, Grand Theft By Employee	Embezzlement, Grand Theft By Employee			
Miscellaneous Investigation	Miscellaneous Investigation			
Suspicious Occurrence	Suspicious Occurrence			
	Casa Clasura			

CLEANING TEXT – REPLACE VALUE



NUMBER CLEANING



ADDITIONAL INFORMATION

✓ Books

✓ Collect, Combine, and Transform Data Using Power Query in Excel and Power BI (2018) by Gil Raviv

✓ Websites

- √ http://datachant.com
- √ https://blog.crossjoin.co.uk
- √ http://excel-inside.pro
- √ https://www.myonlinetraininghub.com/blog
- √ https://excelchamps.com/power-query

NEXT WEBINAR

Combining Data in Power Query

- Merge Query
- Append

CONCLUSIONS & QUESTIONS











