

## Assignment 1.

### 1. Missing values

**SELECT \***

**FROM [master].[d2].[stg\_Kansas\_city]**

**WHERE [CASE ID] IS NULL OR SOURCE IS NULL OR DEPARTMENT IS NULL OR [WORK GROUP] IS NULL OR**

**TYPE IS NULL OR DETAIL IS NULL OR [CREATION DATE] IS NULL OR [CREATION TIME] IS NULL OR**

**STATUS IS NULL OR [EXCEEDED EST TIMEFRAME] IS NULL OR [CLOSED DATE] IS NULL OR**

**[DAYS TO CLOSE] IS NULL OR [STREET ADDRESS] IS NULL OR [ZIP CODE] IS NULL OR**

**NEIGHBORHOOD IS NULL OR COUNTY IS NULL OR [POLICE DISTRICT] IS NULL OR**

**[PARCEL ID NO] IS NULL OR LATITUDE IS NULL OR LONGITUDE IS NULL OR**

**CATEGORY<sub>1</sub> IS NULL OR CATEGORY<sub>2</sub> IS NULL OR CATEGORY<sub>3</sub> IS NULL;**

The screenshot shows the Microsoft SQL Server Enterprise Manager interface. The central pane displays a SQL query that selects all columns from the [master].[d2].[stg\_Kansas\_city] table where any of the following fields are NULL: CASE ID, SOURCE, DEPARTMENT, WORK GROUP, TYPE, DETAIL, CREATION DATE, CREATION TIME, STATUS, EXCEEDED EST TIMEFRAME, CLOSED DATE, DAYS TO CLOSE, STREET ADDRESS, ZIP CODE, NEIGHBORHOOD, COUNTY, POLICE DISTRICT, PARCEL ID NO, LATITUDE, LONGITUDE, CATEGORY<sub>1</sub>, CATEGORY<sub>2</sub>, or CATEGORY<sub>3</sub>. The Results pane below the query shows 16 rows of data with columns: CASE ID, SOURCE, DEPARTMENT, WORK GROUP, TYPE, DETAIL, CREATION DATE, and CREATION TIME. The Properties pane on the right shows the current connection parameters for the 'localhost (d2)' connection.

CASE ID	SOURCE	DEPARTMENT	WORK GROUP	TYPE	DETAIL	CREATION DATE	CREATION TIME
2020054721	WEB	Parks and Rec	Parks and Rec-Central Region-	Park Maintenance	Central	04/18/2020	
2019184705	WEB	NHS	NHS-Solid Waste-	Recycling	Missed by City	10/25/2019	
2019184590	WEB	Parks and Rec	Parks and Rec-Landscape Services-Forestry	Trimming	Tree Limbs	10/25/2019	
2015094486	PHONE	NHS	NHS-Solid Waste-Administration	Services	Service Issue / Problem	08/07/2015	
2020021142	PHONE	Water Services	Water Services-Meter and Field Services-	Leak	Meter	03/11/2020	
2019202956	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/24/2019	
2019202957	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/24/2019	
2019202958	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/25/2019	
2019202959	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/25/2019	
2009066472	PHONE	NHS	NHS-Solid Waste-	OB	All	03/18/2009	
2020134616	PHONE	NHS	NHS-Solid Waste-	Dead Animal	Curb	10/12/2020	
2009090321	PHONE	Fire	Fire-	Fire	All	04/13/2009	
2019202977	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/25/2019	
2019202978	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/25/2019	
2019202979	WEB	Public Works	Public Works-Street and Traffic-Streetlights	Street Light	5+ Lights Out	12/25/2019	
2020159177	EMAIL	NHS	NHS-Solid Waste-Abatement Referral	Dumplings	Solid Waste Referral	12/29/2020	

### Using Alteryx

The screenshot shows the Alteryx interface. The top pane displays a workflow with several tools: Input Data, Data Cleansing, Filter, Formula, Sample, Select, Sort, Join, Union, Text To Columns, Summarize, and Comment. The bottom pane shows a data preview for the 'Construction' field, displaying 1 record with 2 fields (COUNTY, Ditch) and 4,519 bytes. The Results pane on the right shows the input data for the 'Construction' field, displaying 2 of 2 fields (COUNTY, Ditch) and 1 record with 2 fields (COUNTY, Ditch) and 4,519 bytes.

COUNTY	Ditch
Clay	Platte County

The screenshot shows the Alteryx Designer interface. On the left, the 'Browse (13) - Configuration' pane displays a data profile for 'Individual\_Project[1].yxmd'. The profile shows 1 record displayed, 2 fields, and 4,519 bytes. The data is categorized into 'CATEGORY2' (64.1%, 10 Pets, Buildings) and 'CREATION DATE' (0.0%, 5,229 06/24/2019, 06/24/2019). On the right, the workflow diagram shows a sequence of tools: '311\_CallCenterSeviceRequests\_Ka', 'In-Database', 'Filter', 'Join', 'Text To Columns', 'Summarize', and 'Comment'. The 'Filter' tool is configured with the condition '[DAYS TO CLOSE] >= "0"'. The 'Summarize' tool is configured with the output 'CREATED BY = "damg7370"'. The 'Comment' tool is configured with the output 'Created\_Date\_Time = DateTimeNow()'. Below the workflow, the 'Results - Browse (13) - Input' pane shows a table with 2 of 2 fields and 1 record displayed, 4,519 bytes. The table has columns 'Record', 'Desc', and 'Report'. The data is as follows:

Record	Desc	Report
1	String	Table - View Browse Tool Report Tab

## 2. Data types

### Inconsistent data types

### Using Alteryx - Before using auto-field functionality:

The screenshot shows the Alteryx Designer interface. On the left, the 'Browse (4) - Configuration' pane displays a data profile for 'Individual\_Project[1].yxmd'. The profile shows 1,563,215 records displayed, 24 fields, and 155 MB. The data is categorized into 'CASE ID' (2007002634, 2007006846, 2007007334, 2007007437, 2007007461, 995 more >), 'SOURCE' (PHONE, 1,204,236; WEB, 211,721; EMAIL, 80,585; SYS, 19,226; INSPE, 14,690; 16 more >), and 'DEPARTMENT' (NHS, 783,094; Public Works, 353,787; Water Services, 216,852; Parks and Rec, 87,954). On the right, the workflow diagram shows a sequence of tools: '311\_CallCenterSeviceRequests\_Ka', 'In-Database', 'Filter', 'Join', 'Text To Columns', 'Summarize', and 'Comment'. The 'Filter' tool is configured with the condition '[DAYS TO CLOSE] >= "0"'. The 'Summarize' tool is configured with the output 'CREATED BY = "damg7370"'. The 'Comment' tool is configured with the output 'Created\_Date\_Time = DateTimeNow()'. Below the workflow, the 'Results - Browse (4) - Input' pane shows a table with 24 of 24 fields and 1 record displayed, 4,519 bytes. The table has columns 'Record', 'Name', 'Type', 'Size', 'Source', and 'Description'. The data is as follows:

Record	Name	Type	Size	Source	Description
1	CASE ID	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
2	SOURCE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
3	DEPARTMENT	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
4	WORK GROUP	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
5	TYPE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
6	DETAIL	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
7	CREATION DATE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
8	CREATION TIME	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
9	STATUS	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
10	EXCEEDED EST TIMEFRAME	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
11	CLOSED DATE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
12	DAYS TO CLOSE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
13	STREET ADDRESS	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
14	ZIP CODE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
15	NEIGHBORHOOD	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
16	COUNTY	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
17	POLICE DISTRICT	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
18	PARCEL ID NO	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
19	LATITUDE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	
20	LONGITUDE	V_WString	254	File: C:\Users\91961\Downloads\311_CallCenterS...	

### After using auto-field functionality: the data type changes

The screenshot shows the Alteryx Designer interface. On the left, the 'Browse (36) - Configuration' pane displays a profile for 'CASE ID' with 1,563,215 records and 23 fields. Below this, the 'SOURCE' and 'DEPARTMENT' sections show bar charts for various categories. The main workspace displays a workflow with a 'Created\_Date' field set to 'Date:TimeNow()' and a 'Created\_By' field set to 'dmg7370\_week'. The 'Results - Browse (36) - Input' pane shows a table with 23 fields and 11 records.

Record	Name	Type	Size	Source	Description
1	CASE ID	Int32	4	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
2	SOURCE	String	5	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
3	DEPARTMENT	V_String	35	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
4	WORK GROUP	V_String	50	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
5	TYPE	V_String	48	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
6	DETAIL	V_String	48	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
7	CREATION DATE	String	10	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
8	CREATION TIME	String	8	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
9	STATUS	String	5	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
10	EXCEEDED EST TIMEFRAME	String	1	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	
11	CLOSED DATE	String	10	File: C:\Users\aryani\Downloads\Yuga\311_CallCe...	

### 3. Unrealistic Values

The Days to Close column consists of negative values which are unrealistic values.

The screenshot shows the Alteryx Designer interface. On the left, the 'Browse (26) - Configuration' pane displays a profile for 'CASE ID' with 29 records and 23 fields. Below this, the 'SOURCE' and 'DEPARTMENT' sections show bar charts for various categories. The main workspace displays a workflow with a '311\_CallCenterServiceRequests\_KansasCity\_2007-March2021.csv' input. The 'Results - Browse (26) - Input' pane shows a table with 23 fields and 11 records. The 'DAYS TO CLOSE' column contains negative values, which are unrealistic.

Record	T TIMEFRAME	CLOSED DATE	DAYS TO CLOSE	STREET ADDRESS	ZIP CODE	NEIGHBORHOOD	COUNTY	POL
1		03/18/2013	-1.0	4000 N Kimball Dr	64161.0	Birmingham Bottoms	Clay	Sho
2		02/09/2015	-21.0	7906 E 117TH PL	64134.0	Kirkside	Jackson	Sou
3		02/09/2015	-19.0	5212 E 40th Ter	64130.0	Vineyard	Jackson	East
4		02/09/2015	-19.0	653 NE 46TH ST	64116.0	Crestview	Clay	Sho
5		02/09/2015	-21.0	3610 N Cleveland Ave	64117.0	Cooley Highlands S	Clay	Sho
6		02/09/2015	-19.0	4701 FARLEY AVE	64129.0	Eastwood Hills East	Jackson	East
7		02/09/2015	-21.0	3212 NE 47th St	64117.0	Sherwood Estates	Clay	Sho
8		02/09/2015	-21.0	3426 Jefferson St	64111.0	Valentine	Jackson	Cent
9		02/09/2015	-21.0	10400 OXFORD AVE	64134.0	Hickman Mills South	Jackson	Sou
10		02/09/2015	-19.0	8900 Troost Ave	64131.0	Boone Hills	Jackson	Met
11		02/09/2015	-21.0	14333 Kelley Rd	64149.0	Longview	Jackson	Sou