Designing Advanced Data Architectures for Business Intelligence

DAMG 7370

Northeastern University

Final Project – Motor Vehicle collisions/Crashes

- Dealing data with three cities
  - New York
  - Chicago
  - Austin

- Crash data is obtained from Department of Transportation which shows information about each traffic crash on city streets. This data is provided on respective city data portals and links are s below.
  - Motor Vehicle Collisions - Crashes | NYC Open Data (cityofnewyork.us)
  - Austin Crash Report Data - Crash Level Records | Open Data | City of Austin Texas
  - Traffic Crashes - Crashes | City of Chicago | Data Portal

- During Phase 2 implementation, A **change request** (Additional requirement) will be provided which you should implement by adjusting your dimensional model and data loads
  - Wherever applicable.

Final Project Details

NYC OpenData

Home Data About Learn Contact Us Sign In

Motor Vehicle Collisions - Crashes Public Safety

The Motor Vehicle Collisions crash table contains details on the crash event. Each row represents a crash event. The Motor Vehicle Collisions data tables contain information from all police reported motor vehicle collisions in NYC. The police report (MV104-AN) is required to be filled out for collisions where someone is injured or killed, or where there...

Last Updated
March 22, 2024

Data Provided By
Police Department (NYPD)

About this Dataset

Updated
March 22, 2024

Data Last Updated
March 22, 2024

Metadata Last Updated
April 19, 2021

Date Created
April 28, 2014

Views
526K

Downloads
162K

Data Provided by
Police Department (NYPD)

Dataset Owner
NYC OpenData

Data Collection

Data Collection	Motor Vehicle Collisions

Dataset Information

Agency	Police Department (NYPD)

Update

Update Frequency	Daily
Automation	Yes
Date Made Public	5/7/2014

Attachments

MVCollisionsDataDictionary_20190813_ERD.xlsx

Show More

What's in this Dataset?

Rows
2.07M

Columns
29

Each row is a
Motor Vehicle Collision

Columns in this Dataset

Column Name	Description	Type
CRASH DATE	Occurrence date of collision	Date & Time
CRASH TIME	Occurrence time of collision	Plain Text
BOROUGH	Borough where collision occurred	Plain Text
ZIP CODE	Postal code of incident occurrence	Plain Text
LATITUDE	Latitude coordinate for Global Coordinate System, WGS 1984, ...	Number
LONGITUDE	Longitude coordinate for Global Coordinate System, WGS 198...	Number
LOCATION	Latitude , Longitude pair	Location

Show All (29)

Final Project Details

CHICAGO DATA PORTAL Chicago Data Portal

Browse

About Data Related Content

Traffic Crashes - Crashes

Transportation

Crash data shows information about each traffic crash on city streets within the City of Chicago limits and under the jurisdiction of Chicago Police Department (CPD). Data are shown as is from the electronic crash reporting system (E-Crash) at CPD, excluding any personally identifiable information. Records are added to the data portal when a crash...

Read more

Last Updated
March 24, 2024

Data Provided By
City of Chicago

Featured Content Using this Data

Traffic Crashes - Vehicles

External Content

A dataset of vehicles related to these crashes

Traffic Crashes - People

External Content

A dataset of people related to these crashes

Vision Zero Chicago

External Content

VISION ZERO CHICAGO

City of Chicago Vision Zero plan to work with the community to prioritize human life and safety of Chicago's streets.

About this Dataset

Updated
March 24, 2024

Data Last Updated	Metadata Last Updated
March 24, 2024	December 13, 2023

Date Created
October 19, 2017

Metadata

Changes and Other Historical Information Useful to Understanding This Dataset	http://dev.cityofchicago.org/open%20data/data%20portal/2020/07/21/traffic-crash-data-source.html | http://dev.cityofchicago.org/open%20data/data%20portal/2020/02/11/traffic-crash-rd-numbers.html
Data Owner	Chicago Police Department
Time Period	2015 to present (All police districts September 2017 - present)
Frequency	Daily

What's in this Dataset?

Rows	Columns	Each row is a
817K	48	Traffic Crash

Columns in this Dataset

Column Name	Description	Type
CRASH_RECORD_ID	This number can be used to link to the same crash in the Vehi...	Plain Text
CRASH_DATE_EST_I	Crash date estimated by desk officer or reporting party (only ...	Plain Text
CRASH_DATE	Date and time of crash as entered by the reporting officer	Date & Time
POSTED_SPEED_LIMIT	Posted speed limit, as determined by reporting officer	Number
TRAFFIC_CONTROL_DEVICE	Traffic control device present at crash location, as determine...	Plain Text
DEVICE_CONDITION	Condition of traffic control device, as determined by reporting...	Plain Text
WEATHER_CONDITION	Weather condition at time of crash, as determined by reportin...	Plain Text

Show All (48)

Final Project Details

data.austintexas.gov
the official City of Austin open data portal

Data | About | User Resources | Contact Us | Sign In

Austin Crash Report Data - Crash Level Records

Transportation and Mobility

Crash data is obtained from the Texas Department of Transportation (TXDOT) Crash Record Information System (CRIS) database, which is populated by reports submitted by Texas Peace Officers throughout the state, including Austin Police Department (APD), and maintained by TXDOT.…

Read more

Last Updated
March 24, 2024

Data Provided By
City of Austin, Texas - data.austintexas.gov

About this Dataset

Updated
March 24, 2024

Data Last Updated
March 24, 2024

Metadata Last Updated
December 18, 2023

Date Created
July 30, 2019

Views
4,002

Downloads
2,285

Data Provided by
City of Austin, Texas - data.austintexas.gov

Dataset Owner
transportation.data@austintexas.gov

Publishing Information

Update Frequency	Daily

Ownership

Program Name	Vision Zero
Department name	Austin Transportation & Public Works

Attachments

dataset_cover_photo

Topics

Category	Transportation and Mobility
Tags	crash, vision zero

What's in this Dataset?

Rows	Columns	Each row is a
148K	**54**	**Crash**

Columns in this Dataset

Column Name	Description	Type
crash_id	TxDOT C.R.I.S. system-generated unique identifying number f…	Number #
crash_fatal_fl	Fatal Crash Identifier - Indicates that the crash involved one o…	Plain Text T
crash_date	Crash Date	Date & Time
crash_time	Crash Time - Time crash occurred	Plain Text T
case_id	Case ID	Plain Text T
rpt_latitude	Reported Latitude	Number #
rpt_longitude	Reported Longitude	Number #

Show All (54)

Final Project Details

NYC OpenData

Home Data About Learn Contact Us Sign In

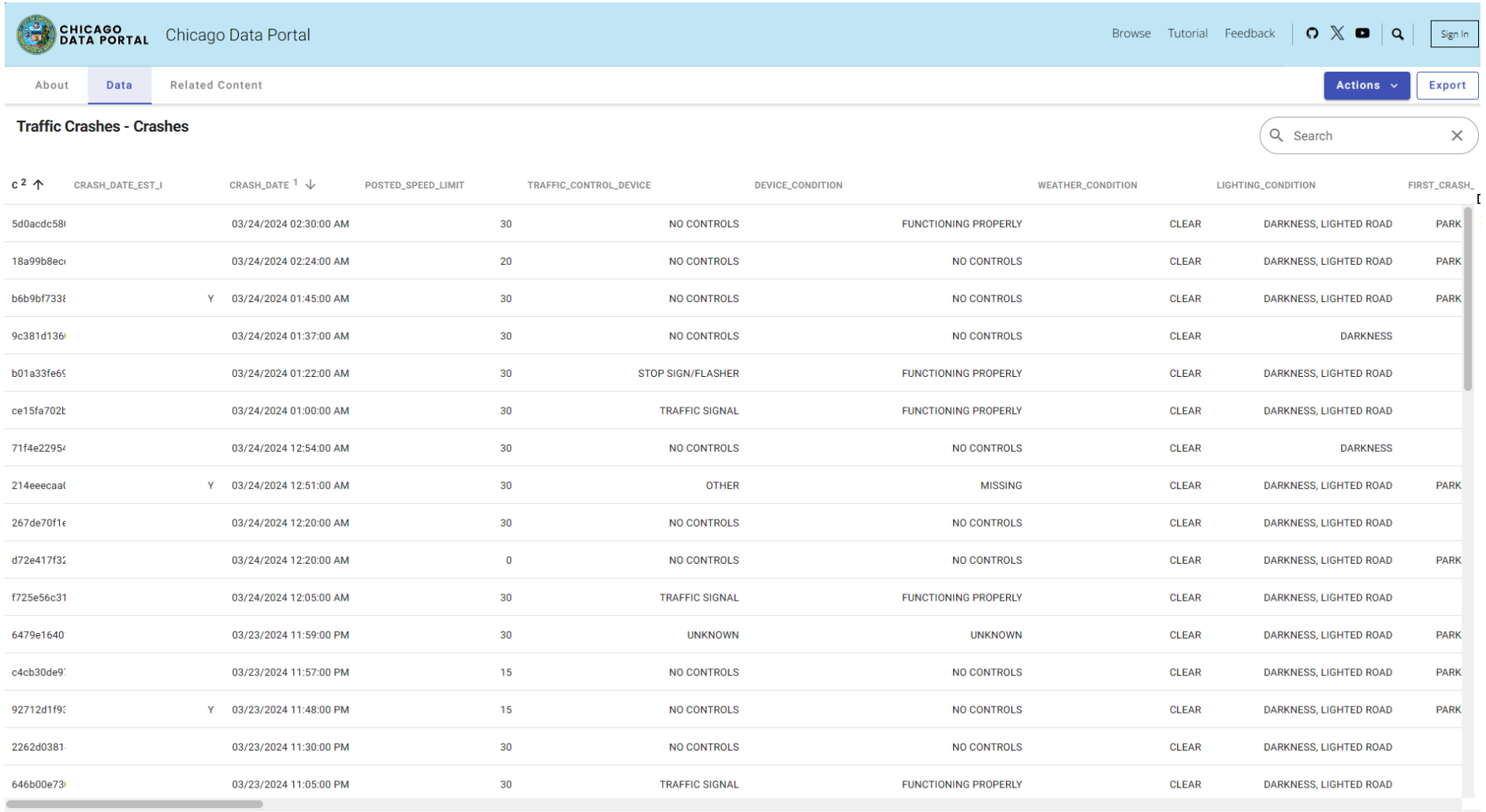
About Data Related Content Actions Export

Motor Vehicle Collisions - Crashes

Search

CRASH DATE	CRASH TIME	BOROUGH	ZIP CODE	LATITUDE	LONGITUDE	LOCATION	ON STREET NAME	CROSS STREET NAME	OFF STREET NAM
09/11/2021	2:39						WHITESTONE EXPRESSWAY	20 AVENUE	
03/26/2022	11:45						QUEENSBORO BRIDGE UPPER		
06/29/2022	6:55						THROGS NECK BRIDGE		
09/11/2021	9:35	BROOKLYN	11208	40.667202	-73.8665	(40.667202°, -73.8665			1211 LORING A
12/14/2021	8:13	BROOKLYN	11233	40.683304	-73.917274	(40.683304°, -73.9172	SARATOGA AVENUE	DECATUR STREET	
04/14/2021	12:47						MAJOR DEEGAN EXPRESSWAY RAMP		
12/14/2021	17:05			40.709183	-73.956825	(40.709183°, -73.9568	BROOKLYN QUEENS EXPRESSWAY		
12/14/2021	8:17	BRONX	10475	40.86816	-73.83148	(40.86816°, -73.83148			344 BAYCHEST
12/14/2021	21:10	BROOKLYN	11207	40.67172	-73.8971	(40.67172°, -73.8971°			2047 PITKIN AV
12/14/2021	14:58	MANHATTAN	10017	40.75144	-73.97397	(40.75144°, -73.97397	3 AVENUE	EAST 43 STREET	
12/13/2021	0:34			40.701275	-73.88887	(40.701275°, -73.8888	MYRTLE AVENUE		
12/14/2021	16:50	QUEENS	11413	40.675884	-73.75577	(40.675884°, -73.7557	SPRINGFIELD BOULEVARD	EAST GATE PLAZA	
12/14/2021	8:30						broadway	west 80 street -west 81 street	
12/14/2021	0:59			40.59662	-74.00231	(40.59662°, -74.00231	BELT PARKWAY		
12/14/2021	23:10	QUEENS	11434	40.66684	-73.78941	(40.66684°, -73.78941	NORTH CONDUIT AVENUE	150 STREET	

Final Project Details



c²	CRASH_DATE_EST_I	CRASH_DATE ¹	POSTED_SPEED_LIMIT	TRAFFIC_CONTROL_DEVICE	DEVICE_CONDITION	WEATHER_CONDITION	LIGHTING_CONDITION	FIRST_CRASH_
5d0acdc58		03/24/2024 02:30:00 AM	30	NO CONTROLS	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD	PARK
18a99b8ec		03/24/2024 02:24:00 AM	20	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	PARK
b6b9bf733	Y	03/24/2024 01:45:00 AM	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	PARK
9c381d136		03/24/2024 01:37:00 AM	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS	
b01a33fe6		03/24/2024 01:22:00 AM	30	STOP SIGN/FLASHER	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD	
ce15fa702		03/24/2024 01:00:00 AM	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD	
71f4e2295		03/24/2024 12:54:00 AM	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS	
214eeecaa	Y	03/24/2024 12:51:00 AM	30	OTHER	MISSING	CLEAR	DARKNESS, LIGHTED ROAD	PARK
267de70f1		03/24/2024 12:20:00 AM	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	
d72e417f3		03/24/2024 12:20:00 AM	0	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	PARK
f725e56c31		03/24/2024 12:05:00 AM	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD	
6479e1640		03/23/2024 11:59:00 PM	30	UNKNOWN	UNKNOWN	CLEAR	DARKNESS, LIGHTED ROAD	PARK
c4cb30de9		03/23/2024 11:57:00 PM	15	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	PARK
92712d1f9	Y	03/23/2024 11:48:00 PM	15	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	PARK
2262d0381		03/23/2024 11:30:00 PM	30	NO CONTROLS	NO CONTROLS	CLEAR	DARKNESS, LIGHTED ROAD	
646b00e73		03/23/2024 11:05:00 PM	30	TRAFFIC SIGNAL	FUNCTIONING PROPERLY	CLEAR	DARKNESS, LIGHTED ROAD	

Final Project Details

data.austintexas.gov
the official City of Austin open data portal

Data | About | User Resources | Contact Us | Sign In

About | Data | Related Content | Actions | Export

Austin Crash Report Data - Crash Level Records

Search

crash...	crash...	crash...	crash...	case_id	rpt_la...	rpt_lo...	rpt_bl...	rpt_st...	rpt_st...	rpt_st...	crash...	road_...	latitude	longit...	street...	street...	street...	street...	crash...	sus_s...	nonin...	poss_...	non_i
14,092,402	N	2014 Oct 2(	11:09:00	143010654					5403 OVERI	DR	25	N			5403 OVERI		PECAN SPR		5	0	0	0	
13,797,332	N	2014 Apr 0(	14:09:00	140991015			8000		E US 290 H'		60	N	30.3276207	-97.662680	US0290		N/A		5	0	0	0	
13,756,880	N	2014 Mar 2	13:47:00	140840887			1500		ANDERSON		-1	N	30.3516811	-97.721541	ANDERSON		N/A		5	0	0	0	
13,802,618	N	2014 Apr 1'	18:40:00	141011587					NEWMAN D	DR	30	N			NEWMAN D		W W 7TH S'		2	0	1	0	
13,784,770	N	2014 Apr 1:	23:22:00	141022046			15000	S	N IH 35 SB	HWY	60	N	30.3227743	-97.707590	IH0035		US0290		5	0	0	0	
13,786,430	N	2014 Apr 0(	00:11:00	140940020			3400	N	IH 35 NB		55	N	30.296133(	-97.718831	IH0035		N/A		5	0	0	0	
13,792,462	N	2014 Apr 0(	08:56:00	140960602					N US 183 H		-1	N	30.3525363	-97.713892	US0183		ANDERSON		5	0	0	0	
13,790,260	N	2014 Apr 0:	08:05:00	140920463			9911	N	BRODIE LN	LN	-1	N	30.18452	-97.84858	N BRODIE L		N/A		2	0	1	0	
13,803,239	N	2014 Apr 2:	21:27:00	141121823			401		LITTLE TEX	LN	35	N	30.2010563	-97.772562	LITTLE TEX	401	N/A		5	0	0	0	
13,791,012	N	2014 Apr 1(	23:25:00	141061973			100	W	ANDERSON	LN	-1	N	30.3439081	-97.705169	US0183	100	IH0035		5	0	0	0	
13,800,841	N	2014 Apr 2:	22:44:00	14-09984			7700		JOHNNY M	RD	-1	N	30.3118484	-97.637548	JOHNNY M		N/A		3	0	0	1	
13,762,289	N	2014 Mar 2	23:40:00	140872255					NOT REPOF		55	N	30.4399191	-97.669320	IH0035		WELLS BRA		0	0	0	0	
13,797,139	N	2014 Apr 0(	09:18:00	140990555			5700		CARRY BAC	LN	30	N	30.3345121	-97.805023	CARRY BAC	6179	N/A		0	0	0	0	
13,790,948	N	2014 Apr 0(	15:20:00	140960985			9100		N IH 35 SVF		55	N	30.357847(	-97.688301	IH0035		N/A		3	0	0	9	
13,781,031	N	2014 Mar 3	08:18:00	140890700					OAKBROOK	DR	30	N	30.3909247	-97.682235	SL0275	11900	FM0734		3	0	0	2	
13,767,511	N	2014 Apr 0:	20:42:00	140921852			12400	N	MOPAC EXF		-1	N	30.4210061	-97.703247	SL0001		N/A		3	0	0	1	

Final Project Details

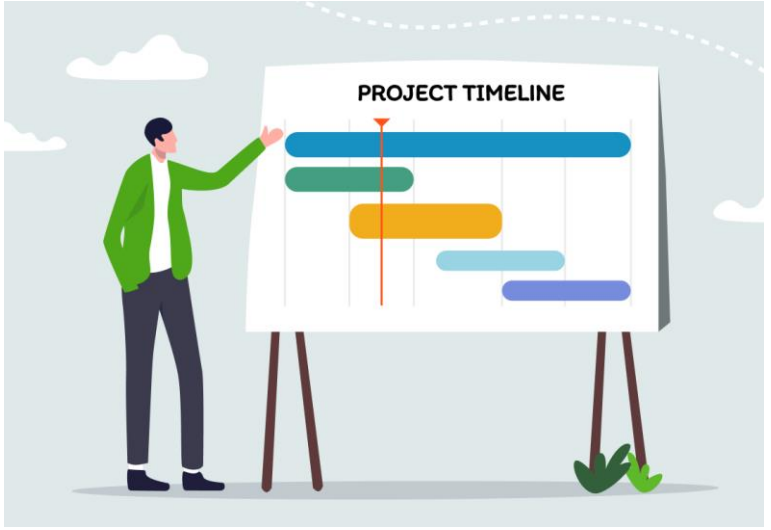
- How many accidents occurred in NYC, Austin and Chicago?
  - Use your ideas on how best to present these values on the dashboard
- Which areas in the 3 cities had the greatest number of accidents?
  - top 3 areas in each city
- How many accidents resulted in just injuries?
  - this report need to be generated at 2 levels, 1 -> overall, 3 -> by city
- How often are pedestrians involved in accidents?
  - this report need to be generated at 2 levels, 1 -> overall, 3 -> by city
- When do most accidents happen?
  - seasonality report

Final Project Details

- How many motorists are injured or killed in accidents?
  - this report need to be generated at 2 levels, 1 -> overall, 3 -> by city
- Which top 5 areas in 3 cities have the most fatal number of accidents?
- Time based analysis of accidents
  - Time of the day, day of the week, weekdays or weekends.
- Fatality analysis
  - Are pedestrians killed more often than road users?
- What are the most common factors involved in accidents?

Final Project Details

- submission timeline
  - Part 1
    - 31st March EOD
  - Part 2
    - 7th April EOD
  - Part 3
    - 14th April EOD



Final Project Details

- Part 1 (all 3 data sets)
  - Data profiling Alteryx / ydata profile
    - Analysis document
  - Data staging (Staging tables)
    - Use talend for ETL jobs
      - Incorporate all standard practices discussed
    - Azure SQL server / MySql / SQL Server
  - Dimensional model (Target tables)
    - Facts and Dimensions
    - Create mapping document
      - Clearly explain the source column name and where it finally maps to target column
        - Stage to Target
        - Document all transformations if any

Final Project Details

- Part 1 (all 3 data sets)
  - Data profiling Alteryx / ydata profile
    - Analysis document
  - Data staging (Staging tables)
    - Use talend for ETL jobs
      - Incorporate all standard practices discussed
    - Azure SQL server / MySql / SQL Server
  - Dimensional model (Target tables)
    - Facts and Dimensions to be created
    - Create mapping document
      - Clearly explain the source column name and where it finally maps to target column
        - Stage to Target
        - Document all transformations if any

- Note
  - All scripts to be uploaded
    - SQL scripts
    - Validation scripts
  - Make sure row counts match to the file rows
  - Take screen shots and upload them
  - This is a team submission
    - One person in a team should submit
  - Upload all documents in 1 zip file
  - Ask for Mapping document template if you need one.
  - Must configure at least one dimension as SCD2
  - Address null values appropriately
  - Maintain Source DIM table and audit columns wherever applicable

Final Project Details

- Part 2 (all 3 data sets)
  - Staging to Integration
    - Using Talend ETL jobs
    - Query dimensional to validate data
    - If any rows rejected explain clearly the reason for rejection
    - Query dimensional data model using SQL for the provided business questions

- **Note** A change request will be provided at the start of the phase 2

Final Project Details

- Part 3
  - Visualizations
    - Tableau and Power BI
      - Note publishing reports to cloud is optional
        - If access issue is resolved, then you can publish reports
    - Upload all screen shots
    - Upload source workbooks

Questions?