Operation 3SL

Slow Speed Saves Lives

Phase 2

System Representation: Design

CIS 9590

Section QMWA

Suhaib Alsahybi	suhaib.alsahybi@baruchmail.cuny.edu
Sage Ashique	abdur.ashique@baruchmail.cuny.edu
John Ortiz	john.ortiz1@baruchmail.cuny.edu
Paridhi Sharma	paridhi.sharma@baruchmail.cuny.edu
Eren Yuceer	ismail.yuceer@baruchmail.cuny.edu

Summary

The Vision Zero mission is to reduce traffic injuries and get traffic fatalities down to zero. The O3SL project aims to support this mission with an interactive dashboard (with a recommendation engine) as well as a report generation feature for the NYPD. The design for the O3SL project takes into account strict stakeholder requirements as well as user experience, culminating in an easy-to-use decision making tool.

The entire flow of data for this project begins at the database level. With partnership and permissions from the NYPD technology department, we will have access to traffic and accident related tables in their database. Using an extract, transform, and load process from their source system, we will pull their data into a temporary, clean it, and then deliver it to a local data warehouse. A scheduled import would run nightly on the production environment in order to maintain a fresh batch of data in the warehouse. During the development phase, the data warehouse would be stored locally, but be moved to a cloud data warehouse before the project goes live. Our selected analytics platform, Tableau, would then directly connect to the data warehouse and import the data for analysis. Our developers would then use tableau to create the interactive dashboards and reports based around the requirements specified by the project stakeholders.

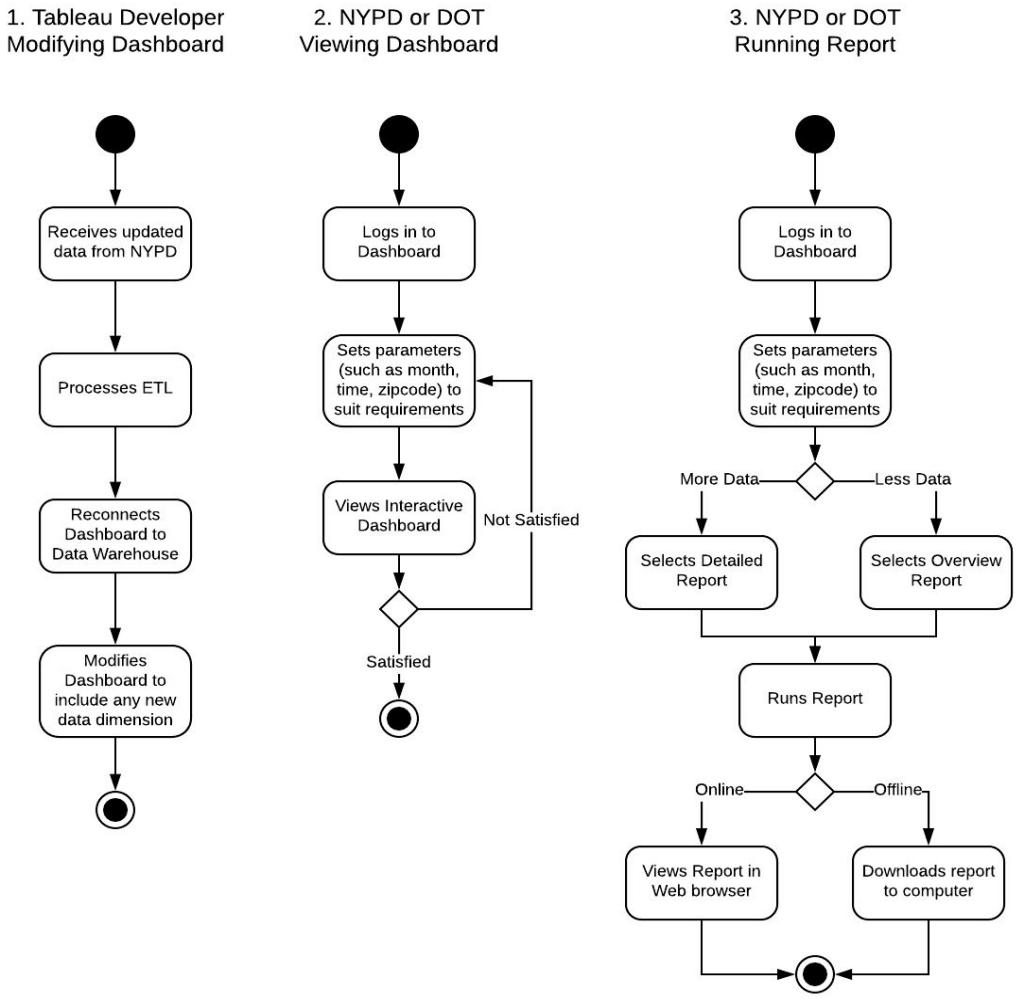
In this document, we will provide details around many of our data processes and how the data will be processed through NYPD to O3SL. The representation of the data flow diagram is provided to explain what the processes do in each phases and explained by the inputs and outputs. Additionally, Use Case diagrams and activity diagrams are shown in the report to emphasize what different types of user, such as O3SL tableau developer, NYPD user and DOT user, do with the tool and the data. While the use cases show the interaction of the users with system, the decision table and tree shows what system enables users to do in the given conditions.

In addition to the above mentioned content, the report contains the process modeling techniques to show how the system transforms data into useful information through visualization. File descriptions, and data types that are used are also categorized and explained.

Activity Diagrams

The three main activities are in the diagrams below. Diagram 1 shows a the Tableau developer when there are changes to be made to the dashboard. These are not changes that can be left to the end user and are not simply selecting attributes. This only occurs when there is a fundamental change in the data being fed into the data warehouse. There might be more types of information and they all have to be vetted to see if they're necessary to the dashboard. We do not want extra data bogging down the system.

The next two diagrams are similar, and from the perspective of the end-user, in this case NYPD or DOT. Once they're logged in, they're greeted by the dashboard and options. They can choose to manipulate the dashboard and see it change live, or select predetermined reports to run.

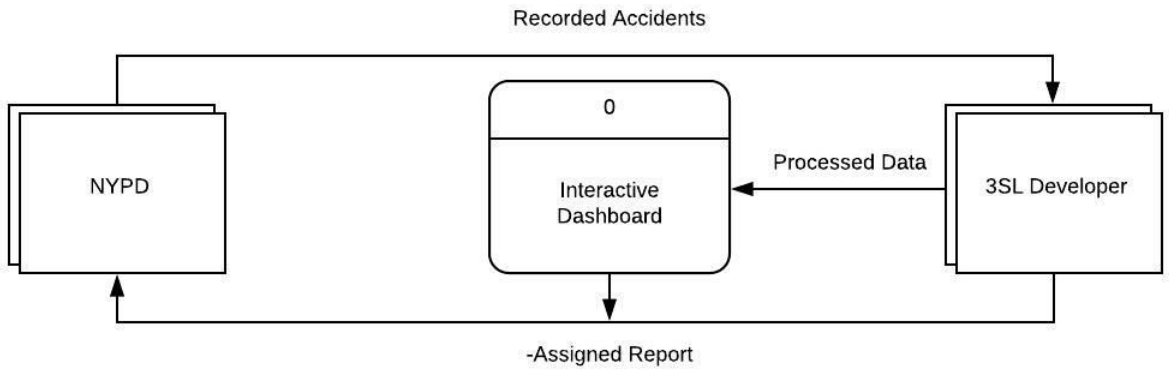


Data Dictionary

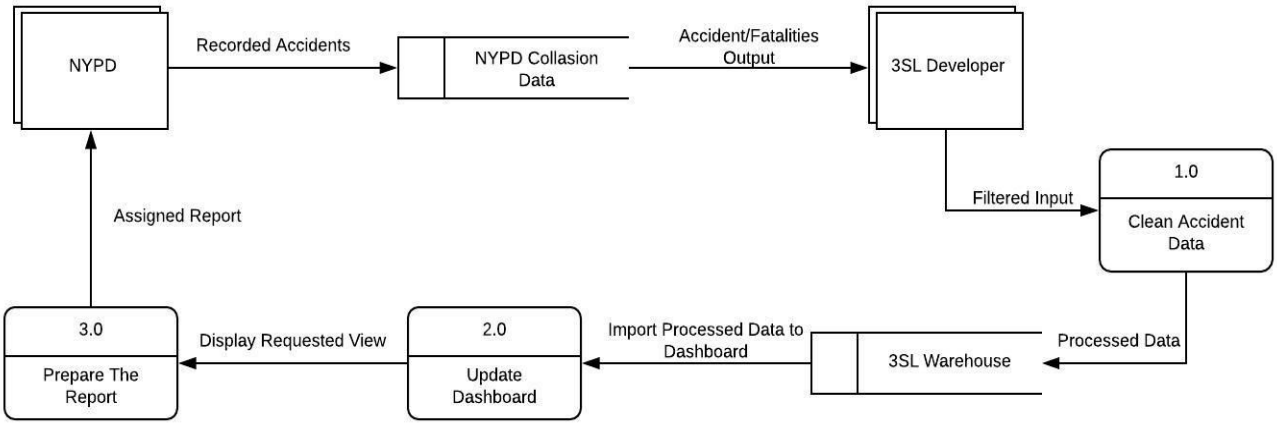
Data Flow Diagrams show how the information system converts input data into output information through a series of processes. The depict how data from external entities (users) flows through the system to output analytical information or simple transaction confirmations. The logical steps described by the series of processes depicted in DFD's are later used to model actual program code.

Data Flows: composition of major flows

The context diagram displays the entire system as single process, in other words Process 0 and shows what data input it receives from external entities which results in output back to the users. The diagram below shows the data flows between the two entities which are NYPD and O3SL Developer.



The Diagram 0 DFD breaks down the general process shown in Process 0. It has one data store O3SL Warehouse and one database NYPD Collision Data. The diagram represents how data is taken from NYPD database, filtered by O3SL developer and cleaned before stored in O3SL Warehouse which is ready to be displayed in Interactive Dashboard. Then, upon requested report from NYPD, the dashboard provides the information requested.



The important point is that dashboard is only a displayer in this case. It only shows data. The data flow back to external entity NYPD only if the reports is taken and downloaded.

Data Dictionary provides specific details on the characteristics of data flows. It is describes the data flow diagrams in a more complete picture that is translatable to a physical implementation. Below we provided the key data dictionary documents describing multiple aspects of the diagrams which plays major role in the system:

Data Flow	
Label	Recorded Accidents
Origin Type	Entity
Origin Label	NYPD
Destination Type	Data Store
Destination Label	NYPD Collision Data
Description	Data on the accidents recorded by the NYPD directly
Attributes	Date, Time, Borough, Contributing Factor Vehicle 1-5, Cross Street Name,
	On Street Name, Off Street Name, Zip Code, Latitude,Longtitude,
	Number of Cyclist Killed, Number Of Pedestrians Killed, Numbrer of Persons Killed,
	Number of Motorists Killed, Number of Cyclists Injured,
	Number of Motorists Injured,Number of Pedestrians Injured
	Number of Persons injured

Data Flow	
Label	Processed Data
Origin Type	Process
Origin Label	Clean Accident Data
Destination Type	Data Warehouse
Destination Label	3SL Warehouse
Description	Data on accident records cleaned according to the Dashboard requirements
Attributes	Date,Time, Borough,Unsafe Speed, Aggressive Driving/Road Rage
	Injuries, Fatalities, Latitude, Longtitude

Major Processes

There are three main major processes in the system: 1.0 - Clean Accident Data, 2.0 - Update Dashboard, 3.0 - Prepare the report. Please see below for the detailed description of processes and structured english to show what does process do basically:

Process Description	
Label	1.0 - Clean Accident Data
Input Data	Filtered Accident Input
Output Data	Processed Data
Process Description	For each Filtered Accident Input
	If input has Vehicle Type Columns:
	Drop Vehicle Type
	If Input has number of persons killed, pedestrians killed,
	motorist killed:
	Combine number under the " Fatalities"
	And
	Drop the column : Number of Persons Killed, Pedestrians Killed,Cyclist Killed,
	Motorists Killed
	If input has numbr of persons injured, pedestrians injured, cyclist injured
	motirist injured:
	Combine number under the "Injuries"
	If columns has Uppercase:
	Convert it to lowercase
	If columns has space :
	Convert it to undrscore
	If zipcode and borough has numerical or ordinal value:
	Convert it to Categorical Value
	If input has 0 and Null Values in lattitude and longtitude:
	Remove lat/long with 0 and Null Values
	Drop Columns contains contributing_factor_vehicle
	Else Pass

Process Description	
Label	2.0 - Update Dashboard
Input Data	Imported Processed Data
Output Data	Displayed Requested View
Process Description	For each processed data
	If the processed data == previous processed data
	Import processed data to dashboard
	Output in Requested View
	Else
	Pass

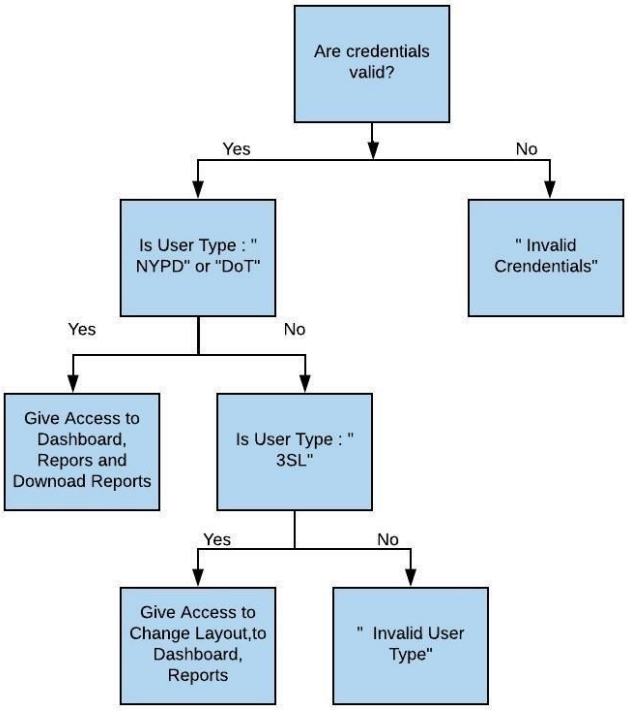
Process Description	
Label	3.0 - Prepare the Report
Input Data	Displayed Requested View
Output Data	Assigned Report
Process Description	If user choose request information from data warehouse:
	Bring requested view with requested data to user with assigned report type
	Else
	Pass

As it can be seen from the process description, processes are only can be applied O3SL developer since O3SL is the party that which makes the cleaning over NYPD data.

The below decision table illustrates the test cases and scenarios on what platform executes/allows according to the user types:

ID	Conditions/Actions	Test Case I	Test Case II	Test Case III	Test Case IV	Test Case V	Test Case VI
Condition I	Valid User ID	F	F	T	T	T	T
Condition II	Valid Password	F	T	F	F	T	T
Condition III	Type of User = "NYPD" or "DoT"	T	T	T	F	T	F
Condition IV	Type of User = "O3SL"	F	F	F	T	F	T
Action I	Show Message " Invalid Credentials"	Execute	Execute	Execute	Execute		
Action II	Land in Dashboard Page					Execute	Execute
Action III	Give Access to Reports					Execute	Execute
Action IV	Give Access to Download Reports					Execute	Execute
Action V	Give Access to Change Layout						Execute

In the given conditions, the validations can be applied to decision tree as shown at below:



Data Stores: composition of data stores

The following table describes our target data dictionary. This dataset will be feed into the tableau environment where a visual analytics dashboard will be generated and displayed to the end user.

Variable	Description	Datatype
date	Date of incident	date
time	Time of incident	time
borough	Borough of incident	catergorical
zip_code	Zip code of incident	catergorical
latitude	latitude coordinate of incident	float
longitude	longitude coordinate of incident	float
on_street_name	Street name within the legal driving area	string (object)
cross_street_name	Intersection crossing	string (object)
off_street_name	Street name Off of legal driving area	string (object)
fatalities	total # of fatalities that took place	integer
injuries	total # of injuries that took place	integer

File Descriptions

After the ETL process from the NYPD database source system is complete, we will store the cleaned data into an O3SL data warehouse (locally during the development process, but provisioned to be hosted on Google Cloud before we go live). The ETL process will run on a nightly basis in order to keep the data live and up to date.

a. **Composition of files - Fields, Lengths, Type of data:** Using Tableau, our developers will import all related data from the O3SL DataWarehouse so that it can be processed according to the user requirements. All fields in Tableau are separated into either Dimensions or Measures.

- Dimensions are qualitative and do not total a sum.
- Measures are numerical values that mathematical functions work on.

Tableau classifies every piece of data into one of four categories, namely - String, Number, Boolean and datetime, and does not include length restrictions for each data type. Below is ae list of the 23 fields selected for import into tableau along with their data types:

#	Field	DataType
1	Unique Key	Dimension: Number (Whole)
2	Date	Dimension: Date
3	Time	Dimension: Date & Time
4	Borough	Dimension: String
5	Contributing Factor Vehicle 1	Dimension: String
6	Contributing Factor Vehicle 2	Dimension: String
7	Contributing Factor Vehicle 3	Dimension: String
8	Contributing Factor Vehicle 4	Dimension: String
9	Contributing Factor Vehicle 5	Dimension: String
10	Cross Street Name	Dimension: String
11	On Street Name	Dimension: String
12	Off Street Name	Dimension: String
13	Zip Code	Dimension: Number (Whole)
14	Latitude	Measure: Number (Decimal)
15	Longitude	Measure: Number (Decimal)
16	Number of Cyclist Killed	Measure: Number (Whole)
17	Number of Motorist Killed	Measure: Number (Whole)
18	Number of Pedestrians Killed	Measure: Number (Whole)
19	Number of Persons Killed	Measure: Number (Whole)

20	Number of Cyclist Injured	Measure: Number (Whole)
21	Number of Motorist Injured	Measure: Number (Whole)
22	Number of Pedestrians Injured	Measure: Number (Whole)
23	Number of Persons Injured	Measure: Number (Whole)

b. **Primary and foreign keys**: The data warehouse in use is of the Star Schema type. The 6 dimensions are Location, Accident, Date, Time, Vehicle, and Accident Fact table. Each of the supporting dimensions has a corresponding surrogate key that line up to the fact table, creating a composite primary key, all using a 1 to many relationship.

- *LocationDimId* is the surrogate key to the Accident fact table originating from the *LocationDimension*
- *DateDimId* is the surrogate key to the Accident fact table originating from the DateDimension
- *TimeDimId* is the surrogate key to the Accident fact table originating from the *TimeDimension*
- *VehicleDimId* is the surrogate key to the Accident fact table originating from the *VehicleDimension*
- *AccidentDimId* is the surrogate key to the Accident fact table originating from the *AccidentDimension*

The data that O3SL is using is extracted from an existing NYPD database, transformed, and loaded into a data warehouse. For that reason, all the data marts within the data warehouse were denormalized, for analytics optimization and performance, during the ETL process when creating the data warehouse.

c. **Attribute and Record Selection:** As part of our data cleaning process, we removed attributes and record types which did not add value to the requirements of this dashboard. By reducing the amount of unnecessary records, we were able to reduce the file size which will help in performance when running analysis queries. As an example, records were removed where the reason for accident were ones such as “driver Illness” or “texting”, as we believe locations for these types of violations are random in nature and will serve no purpose when evaluation high risk locations.
The original data set is available on request as well as directly from the NYPD information technology department.
Our final set for import into our analytics platform:

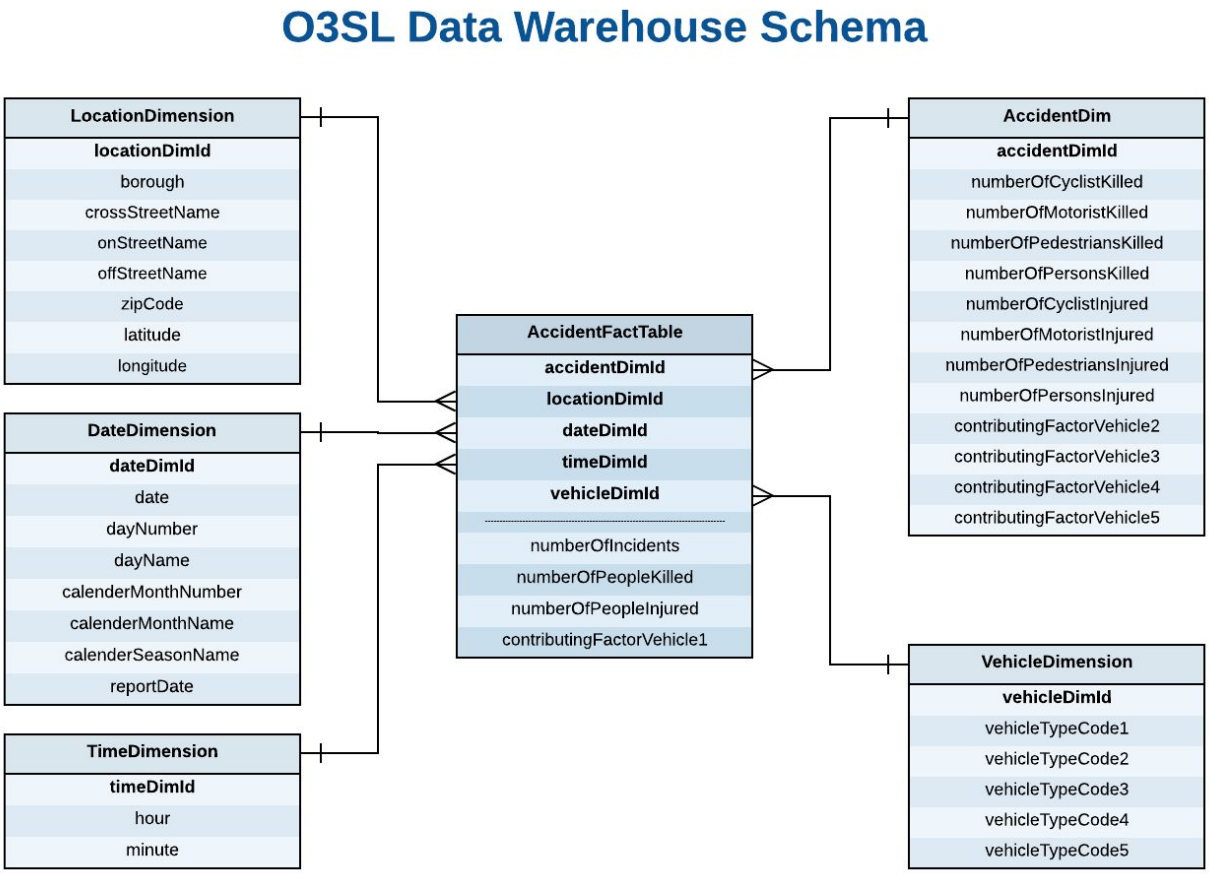
1. Includes only ‘Aggressive Driving’ and ‘Unsafe Speed’ for the Accident reason (Contributing Factor)

2. Does not include Vehicle Type (Code) from our field selections as the vehicle type did not weigh heavily in terms of locations for vehicle accidents.

Our final dataset included 15,074 rows worth of data for vehicle accidents in New York City.

File Relationships

Below is the O3SL Data Warehouse Schema Model prior to import into the analytics platform.

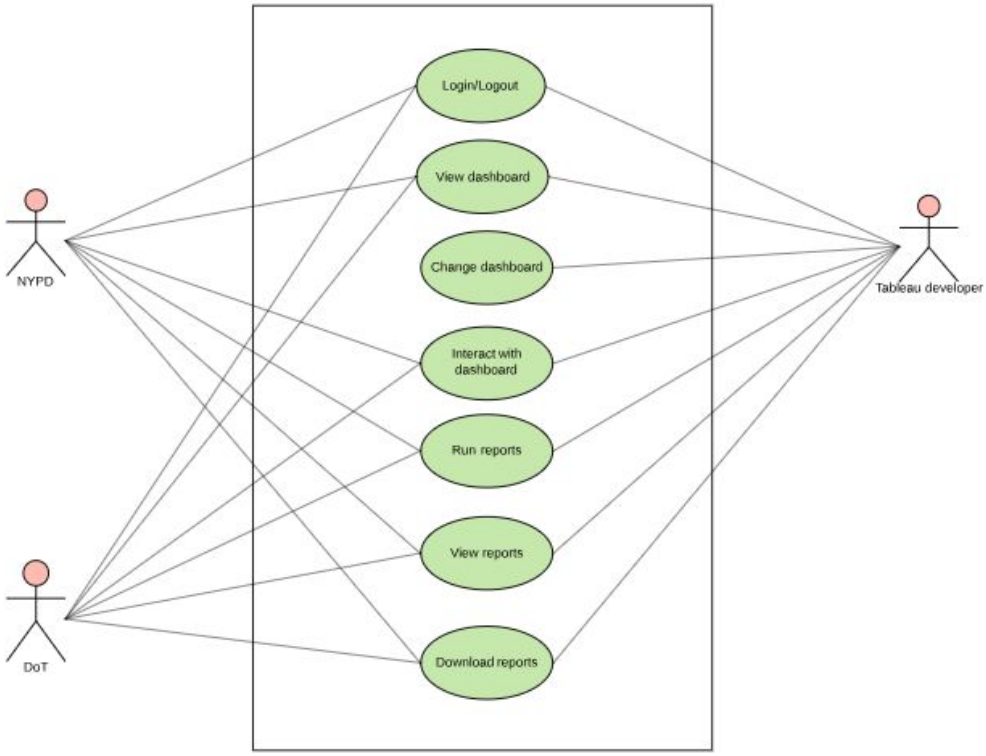


Use Cases

The below use case diagram outlines high level overview of the interaction between the system's users also known as actors with the system.

Actors : NYPD, DoT and Tableau Developer
System: Tableau
Goal: To access dashboard to view and interact with it with help of filters and selections
To run and download reports from Tableau for reporting and decision purposes.

Users(or actors) Interaction with system(tableau)



Task level goal	Actors			Brief
	NYPD	DoT	Tableau Developer	
Login/Logout in the tableau dashboard	✔	✔	✔	In order to access the dashboard,NYPD, DoT,Tableau developer will be required to login into the dashboard shared through a web page or on Tableau server with their authorized email id and password.
View dashboard	✔	✔	✔	After logging in, All the actors can view the dashboard.
Change dashboard	✘	✘	✔	Only Tableau Developer would have authorization to make any modification,integration to the dashboard.
Interact with dashboard	✔	✔	✔	All the actors can interact with dashboard with various selections and filter by time, day, zip code etc depending upon the output required.
Run reports	✔	✔	✔	All the actors would be authorized to run the output reports based on the selection.
View reports	✔	✔	✔	All the actors can view the reports in the tool itself.
Download reports	✔	✔	✔	All the actors can download the report on their local system to view and analyze it.

Risk Register

Id	Date Raised	Risk Description	Likeli-hood	Impact	Severity	Owner	Mitigating Action	Progress	Status
		DESIGN RISK							
1	11/8/18	Reliance on NYPD Data Warehouse and live connection. Risk of DW failure.	Low	High	High	John Ortiz / Eren Yuceer	Contact NYPD technology. Create plans for future on site data backup if this becomes an ongoing issue.		
2	11/8/18	Original requirements by stakeholders may change based on real world situations (i.e. unknown disaster in city may re-prioritize requirements in initiative)	Low	Med	Med	Paridhi Sharma	Speak to Project Stakeholder regarding change in plans as early as possible. Document all changes and communicate with project team.		
3	11/8/18	Incorrect information or unclear guidelines from Stakeholders for dashboard/report requirements	High	High	Med	Suhaib Alsahybi / Sage Ashique	BA's will speak to stakeholders. Question with higher level of detail. Document all responses and communicate with developers.		
4	11/8/18	Low data quality and data points can affect the output of design.	Med	Med	Med	John Ortiz / Eren Yuceer	Developers will discuss with PM additional methods of retrieving required data if possible. Otherwise communicate with Stakeholder for requirement updates.		
5	11/8/18	Lack of acceptance of design proposal by stakeholder	Med	Med	Low	Paridhi Sharma	PM will speak to stakeholders. Question with higher level of detail. Document issues raised by Stakeholder, update requirements and deadlines, and communicate with developers.		
		TIME RISK							
6	11/8/18	Delays due to dependencies on outside organizations (NYPD Data Docs, access to DW) may lead to unrealistic deadlines in contract.	Med	Med	Low	John Ortiz / Eren Yuceer	Contact PM and communicate with Stakeholder and outside organization which are causing delays. Make aware of deadline shifts due to delay. Have all parties sign off.		
7	11/8/18	Reduction in resources (termination of developer, developer calls out) may lead to additional time required on project	Low	High	Low	Paridhi Sharma	Contact stakeholder immediately on resource termination (only). If decrease in resources will affect timeline, document, inform stakeholder, and get sign off on new deadline dates.		
8	11/8/18	Training of staff is not scheduled in a timely manner can affect go live date	Low	Med	Low	Paridhi Sharma	Contact Stakeholder and department heads for which training is delayed. Make all parties aware of potential shift in deadline.		
		BUDGET RISK							
9	11/8/18	Underestimation of budget design based on original time scope. If deadlines based on time are not maintained, additional increase in budget may be required.	Med	Med	Low	Paridhi Sharma	Perform analysis on triple constraint to see if budget can be maintained. If not possible, contact stakeholder to review budget, and if necessary, request budget increase.		

Software Costing based upon Functional Point Analysis Estimation Model

Function Point Worksheet

Measurement paramet	Count		Weighting Factor simple	average	complex	Choice		
# of user inputs	1	X	3	4	6	4	=	4
# of user outputs	3	X	4	5	7	5	=	15
# of user inquiries	87	X	3	4	6	3	=	261
# of files	1	X	7	10	15	15	=	15
# of external interfaces	12	X	5	7	10	10	=	120
						Count-total =		**415**

Rate each factor on a scale of 0 to 5:	0 - No Influence	**1 - Incidental**	2 - Moderate	
	3 - Average	4 - Significant	5 - Essential	
1	Does the system require reliable backup and recovery?			5
2	Are data communications required?			5
3	Are there distributed processing functions?			5
4	Is performance critical?			5
5	Will the system run in an existing, heavily utilized operational environment?			5
6	Does the system require on-line data entry?			2
7	Does the on-line data entry require the input transaction to be built over multiple screens/operations?			4
8	Are the master files updated on-line?			5
9	Are the inputs, outputs, files, or inquiries complex?			4
10	Is the internal processing complex?			4
11	Is the code designed to be reusable?			5
12	Are conversion and installation included in the design?			5
13	Is the system designed for multiple installations in different organizations?			5
14	Is the application designed to facilitate change and ease of use by the user?			5
			sum of Fi =	**64**

Funtion Point Metric = count-total * [.65+.01*sum Fi]
= 535

Cost of Project = $221,000.00
Cost of Software = $413.08 Per Function Poin

Communications Matrix

COMMUNICATION	PURPOSE	MEDIUM	FREQUENCY	AUDIENCE	OWNER	DELIVERABLE
Kickoff Meeting (Project Charter)	Introduce project. Review objectives and goals.	In person / Face-to-face & Blackboard	Once	Project team Project sponsor Stakeholders	Project manager	Document & Powerpoint
Project Team Meetings	Review status of project, Assign work,	Conference call	Weekly	Project team	Project manager	Agenda Project schedule
Technical & Design Meetings	Discuss, review technical & design problems and solutions.	Whatsapp / Conference call	As needed	Technical team Business team	Technical lead Business lead	Design Decisions
Project Status Meetings	Update on work status, upcoming deadlines, any changes to schedule	Whatsapp / Conference call	Weekly	Project team	Project manager	Schedule & Assigned Work
Phase 1 Problem Definition Presentation	Detailed report on project status including progress, costs, and problems.	In person / Face-to-face & Blackboard	Monthly	Project team Project sponsor Stakeholders	Project manager	Project status report Project schedule
Phase 2 Design Paper & Presentation	Update leadership on project status.	In person / Face-to-face & Blackboard	Once	Project team Project sponsor Stakeholders	Project manager	Design Report & Powerpoint
Phase 3 Prototype Paper & Presentation	Detailed report on project status including progress, costs, and problems.	In person / Face-to-face & Blackboard	Once	Project team Project sponsor Stakeholders	Project manager	Prototype Demo, Report, & Presentation
Phase 4 Implementation Paper & Presentation	Update leadership on project status.	In person / Face-to-face & Blackboard	Once	Project team Project sponsor Stakeholders	Project manager	Implementation Report, presentation, support plan

Decision Log

ID	Decision	Description of decision and impact	Source	Owner	Priority	Date Added	Status	Decision Date	Record of decision made
Unique identifier	Brief 3-5 word description of decision needed	Description of the decision and the impact of making or not making the decision	Meeting/conversation/person from which this need for a decision came	Person who is responsible for ensuring this decision is made	Relative priority: 1-High, 2-Medium, 3-Low	Date this decision was added to this list	Status of item: 1-Not started, 2-In progress, 3-Complete	Date this item was complete.	Notes or results on the completion of this item, such as final decision document
2A	Where to store data warehouse	Deciding between using Tableau Servers, Google storage, or Amazon cloud services. We spoke about the pros and cons of each, decided to go with the familiar service and leave room for expansion.	Group Meeting 1	John	Low	9-Nov	Complete	7-Nov	Design phase document
2B	Decide Presentation Responsibilities	Distribute report and presentation duties. This was needed in order to proceed with the project. Good documentation is essential to our grades.	Entire Project Team	Team	Medium	9-Nov	Complete	5-Nov	Design phase meeting log
2C	What attributes are needed	Deciding how records will be filtered out during the ETL process. Choosing which attributes are essential and which are not. How many records are enough to use for the project? This is essential to the design phase as we cannot make these decisions during implementation. The more that is planned, the less time the actual work will take.	John, Sage, Suhaib	John	High	9-Nov	Complete	9-Nov	Design phase document
2D	Figure out Costs	Look back at estimate and see if they were accurate. Barely anything changed. But now we're using the functional point analysis.	Entire Project Team	Team	High	9-Nov	Complete	10-Nov	Design phase document

Group Meeting Log Sheet 1

Date of Meeting: 11/05/2018 Time of Meeting: 7:00PM

Group: Make It Better Recorder: Suhaib Alsahybi

Attending:
- Sage Ashique
- Eren Yuceer
- Paridhi Sharma
- John Ortiz
- Suhaib Alsahybi

Absent:
- None

Excused?

Meeting Minutes

- Assigned responsibilities
- Created Draft Phase II Document
- Created Draft Phase II Slides
- Outlined approach for use case and activity diagrams
- Discussed approach for ERD and data dictionary
- Discussed Survey and questionnaire results, finalizing what features will be in dashboard

Tasks Assigned	**Team Member**	**Delivery Date**
Work on Part D and E	John Ortiz	11/7/2018
Work on Part D and E	Sage Ashique	11/7/2018
Work on Part C	Eren Yuceer	11/7/2018
Work on Part F and G	Paridhi Sharma	11/7/2018
Work on Part B	Suhaib Alsahybi	11/7/2018

Performance Appraisal & Sign-off **Meeting Ending Time:** 9:00 PM

Team member name	Signature	Contribution
John Ortiz		20%
Sage Ashique		20%
Eren Yuceer		20%
Paridhi Sharma		20%
Suhaib Alsahybi		20%

Group Meeting Log Sheet 2

Date of Meeting: 11/09/2018 **Time of Meeting:** 7:00PM

Group: Make It Better **Recorder:** Suhaib Alsahybi

Attending:
- Sage Ashique
- Eren Yuceer
- Paridhi Sharma
- John Ortiz
- Suhaib Alsahybi

Absent: - None Excused?

Meeting Minutes

- Discussed approach to clean dataset file and design dashboard
- Complete final Report
- Did some exploratory data analysis within the group
- Validated that the project as designed will fulfill all of the initial promised functionality.
- Complete final PPT

Tasks Assigned	Team Member	Delivery Date
File Composition, Data cleaning design (Attribute And Record Selection)	John Ortiz	11/9/2018
File Composition, File Relationships, Risk Register	Sage Ashique	11/9/2018
Major Processes	Eren Yuceer	11/9/2018
Use Cases, Meeting, Appraisal,	Paridhi Sharma	11/9/2018
Activity Diagrams, Decision Log, Communications Matrix	Suhaib Alsahybi	11/9/2018

Performance Appraisal & Sign-off **Meeting Ending Time:** 9:00 PM

Team member name	Signature	Contribution
John Ortiz		20%
Sage Ashique		20%
Eren Yuceer		20%
Paridhi Sharma		20%
Suhaib Alsahybi		20%