



# **REAL TIME TRAFFIC ANALYSIS**

**BDM 1034 - Application  
Design for Big Data**

*ThePhoto by PhotoAuthor is licensed under CC BY SA.*



# Group Members

---

Andres C0931978

Manjot C0938094

Johan C0931102

Anju C0935847

Amandeep C0937432

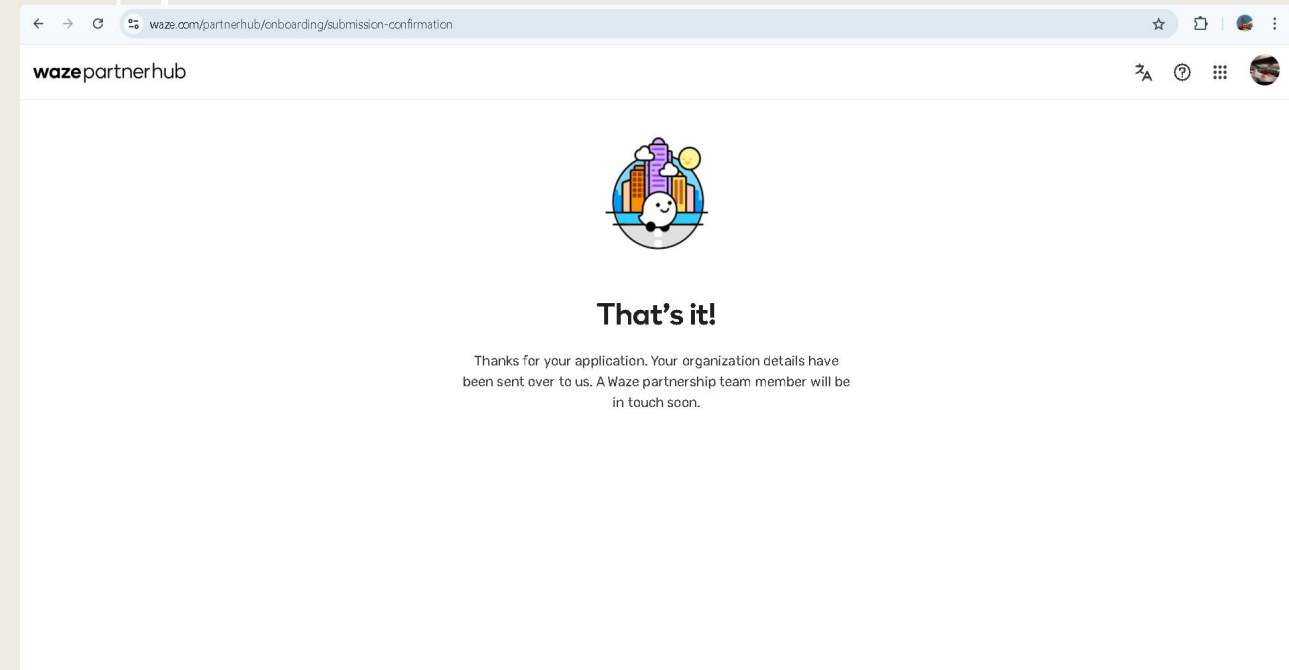
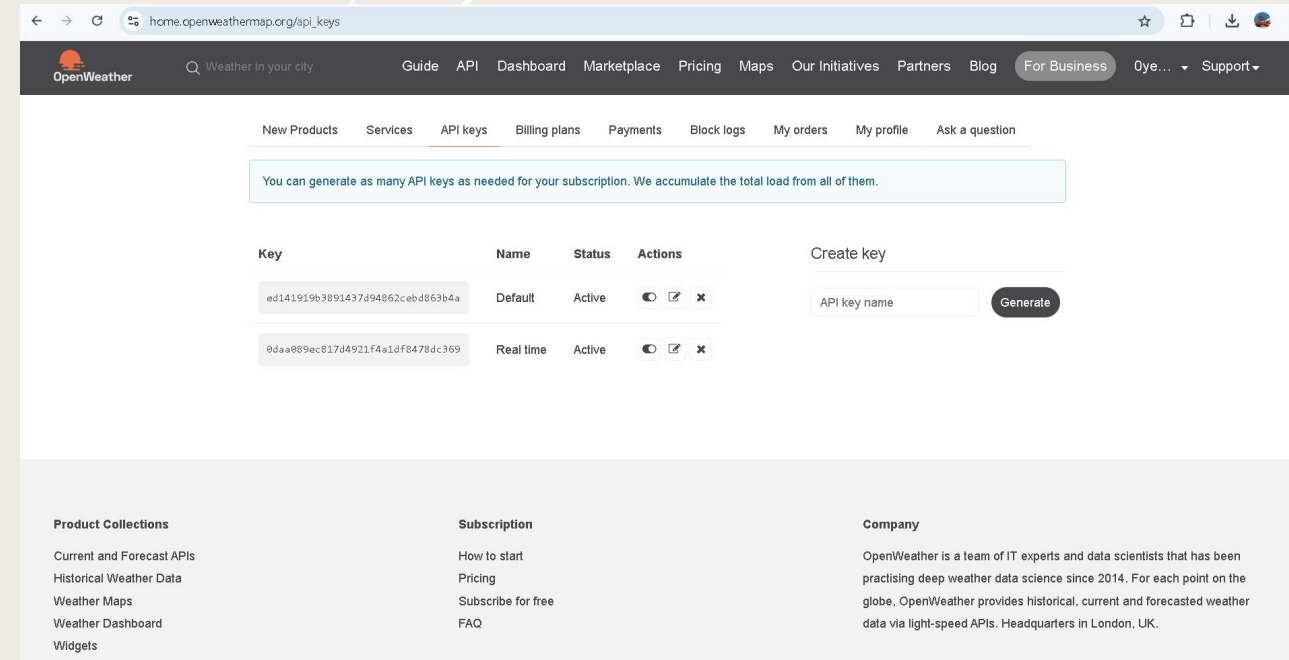
Sarita C0938516

Brian C0940439

Amita C0935607

# Overview of data sources

- Kaggle: Historical traffic data.
- OpenWeather: Real-time weather data affecting traffic.
- Waze: Real-time traffic incidents, road closures



Data Ingestion: Collecting data from various sources like APIs, databases, or files.

Data Transformation: Cleaning the data (removing duplicates, handling missing values), aggregating, or applying necessary computations.

```
def run_spark_pipeline():
    filepath = "C:\\Users\\USUARIO\\Desktop\\projectAD\\US_Accidents_March23.csv"
    data = load_data(filepath)
    cleaned_data = clean_and_transform_data(data)

    # Show the first few rows of the cleaned data
    cleaned_data.show()

# Run the Spark pipeline
run_spark_pipeline()
```

ID	Source	Severity	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance(mi)	Description	Street	
A-729	Source2	3	2016-06-21 10:34:40	2016-06-21 11:04:40	38.0853	-122.2330	1699999999	NULL	NULL	0.0	Right hand should...	Magazine St
A-730	Source2	3	2016-06-21 10:30:16	2016-06-21 11:16:39	37.631813	-122.084167	NULL	NULL	0.0	Accident on I-880...	I-880 N	
A-731	Source2	2	2016-06-21 10:49:14	2016-06-21 11:19:14	37.896564	-122.0707	1699999999	NULL	NULL	0.0	Right lane blocke...	I-680 N
A-732	Source2	3	2016-06-21 10:41:42	2016-06-21 11:11:42	37.334255	-122.032471	NULL	NULL	0.0	#4 & HOV lane...	N De Anza Blvd	
A-733	Source2	2	2016-06-21 10:16:26	2016-06-21 11:04:16	37.250729	-121.9107	1299999999	NULL	NULL	0.0	Right hand should...	Norman Y Mineta Hwy
A-734	Source2	3	2016-06-21 10:31:06	2016-06-21 11:16:26	37.701584	-121.906929	00000004	NULL	NULL	0.0	Accident on I-580...	I-580 E
A-735	Source2	3	2016-06-21 10:17:17	2016-06-21 11:03:50	37.328312	-121.8718	1100000001	NULL	NULL	0.0	#1 lane blocked d...	I-280 S
A-736	Source2	3	2016-06-21 10:51:31	2016-06-21 11:21:31	37.719162	-122.448273	NULL	NULL	0.0	Accident on I-280...	John F Foran Fwy	
A-737	Source2	2	2016-06-21 10:56:00	2016-06-21 11:26:00	37.868114	-122.19593	NULL	NULL	0.0	Right hand should...	CA-24 W	
A-738	Source2	3	2016-06-21 10:57:01	2016-06-21 11:27:01	37.700951	-121.80175	NULL	NULL	0.0	Left lane blocked...	Isabel Ave	
A-739	Source2	2	2016-06-21 11:02:59	2016-06-21 11:32:59	38.221342	-122.713707	NULL	NULL	0.0	Right hand should...	W Sierra Ave	

- Data Storage: Storing the transformed data in a suitable database or data warehouse for further use.
- Data Analysis: Running analytics on the processed data using machine learning models or statistical methods.

```
# Main function for debugging and initial data inspection
def run_debug_pipeline():
    filepath = "C:\\Users\\USUARIO\\Desktop\\projectAD\\US_Accidents_March23.csv"
    data = load_data(filepath)

    if data is None:
        print("Failed to load data. Please check the file and try again.")
        return

    # Show initial data to inspect possible misalignment
    print("Initial raw data:")
    data.show(5)
    data.printSchema()

    # Check initial values of 'Start_Time'
    print("Initial 'Start_Time' values:")
    data.select("Start_Time").show(5, truncate=False)

# Run the debug pipeline to inspect data
run_debug_pipeline()
```

[3]

... Successfully loaded data.

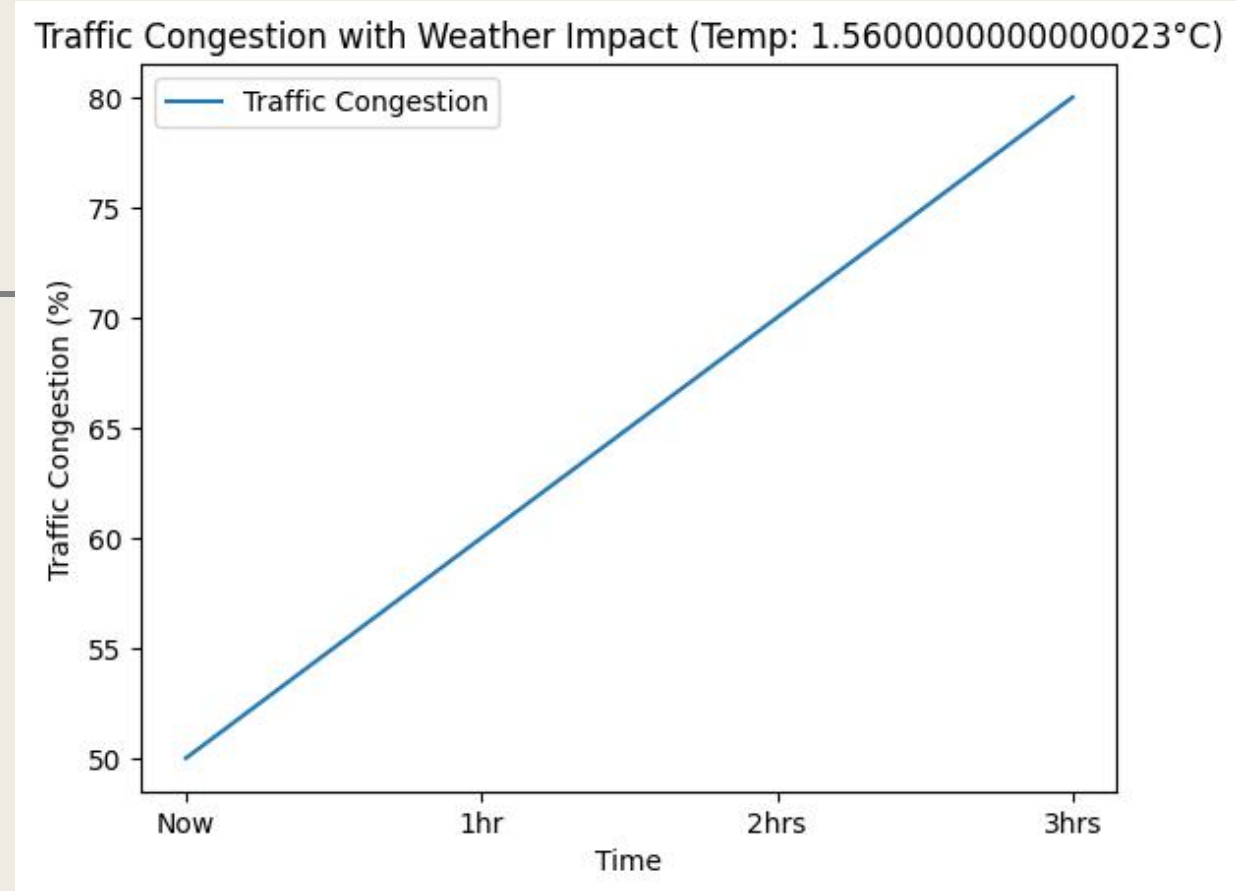
Initial raw data:

```
+---+-----+-----+-----+
| ID|Start_Time|End_Time|          State|
+---+-----+-----+-----+
|A-1|  Source2|      3|2016-02-08 05:46:00|
|A-2|  Source2|      2|2016-02-08 06:07:59|
|A-3|  Source2|      2|2016-02-08 06:49:27|
|A-4|  Source2|      3|2016-02-08 07:23:34|
|A-5|  Source2|      2|2016-02-08 07:39:07|
+---+-----+-----+-----+
```

only showing top 5 rows

# Data Insights

- Correlation between Weather and Traffic: Rainfall and low temperatures increase traffic congestion, leading to longer travel times.
- Emergency Route Optimization: Traffic data (from Waze) combined with weather data can optimize routes for emergency responders by avoiding congested areas.
- Traffic Speed vs. Weather: Inclement weather (rain, snow, low visibility) significantly impacts traffic speed and road conditions, contributing to delays.





# Project Board

Andressanta09 / Projects / Application Design for Big Data project

Q

Type / to search

+

Application Design for Big Data project

Add status update

Prioritized backlog

Status board

Current iteration

Roadmap

Bugs

My items

New view

Filter by keyword or by field

Discard

Save

Title	Status	Assignees	Team	Size	Estimate	Linked pull requests	Iteration
No Priority 9 Estimate: 0							
1 Configuring pipelines and selecting metrics to monitor	In progress	Andressanta09	Data Pipelines				
2 Documentation and analysis of data collection mechanis...	In progress	saritanwar	Data Gathering				Iteration 2
3 Review of the board with progress from the initial sprint	In progress	kaurjot12	Project Board				Iteration 1
4 Analysis and generation of actionable insights	Todo	Oyeaman	Insights				
5 Testing and improving data pipelines	In progress	Jrodriguez8004	Data Pipelines				
6 Completion of data collection and presentation of progr...	Todo	Amita2514	Data Gathering				
7 Final preparation of the project board presentation	In progress	Anju65	Project Board				
8 Preparation and rehearsal of the group presentation	Todo	BrianMartinez10	The whole group				
9 https://github.com/saritanwar/Real-time-traffic-analysis		saritanwar					
+ Add item							

https://github.com/users/Andressanta09/projects/1/views/3

**Thank you (Any  
questions?)**