Cloud–Based Analytics Dashboard

Data
Driven
Highways

Cuadros Rivas, Alejandra Paola – KK5459

Supervisor:Al Nahdi, Amad Salim

Traffic
Accident
Statistics and
Insights

MOTIVATION

Data Driven Highways – Traffic Accident Statistics and Insights

Data Analytics

Context of increasing volumes of data and the need for quick, informed decision-making.

Essential Tool

Centralized, dynamic platform that allows visualization, analysis and interpretation of data

Empower Organizations

Informed decisions and drive business growth

Advantages

Scalability, cost-effectiveness, real-time updates, and simplified collaboration.

PROBLEM SOLVING SITUATION

Challenge

Processing, storing and displaying → seamless, efficient, and accessible.

Cloud Computing

Delivery of computing services over the internet (servers, storage, db, etc.)

Traditional

Short in scalability, cost and the ability to collaborate effectively.

Business

Maintain competitiveness in a rapidly evolving Market

TOOLS AND TECHNOLOGIES

Deployment Environment

Cost Effective
Serverless
Google Cloud Platform
Highly Scalable
Cloud AI
Custom Machine Types
Big Data Analytics
IoT
Internet of Things
API Platform and Ecosystem

Backend

- Python: Logic
- Flask: Framework
- Google cloud Big Query: Datawarehouse

Frontend

- Javascript: Develop front end
- React: Library build user interfaces
- CSS: Styling front end
- Axios: HTTP Request Handling

Development Environment

- Node.js: JS runtime
- Npm: Management tool for JS packages
- Visual Studio Code: IDE

Data Driven Highways Dashboard Blueprints

1. Project Initialization

Objectives: Trend, Severity, Geospatial and Predictive Analysis, and External Factors Correlation

2. Environment Setup GC

- Big Query: Data warehouse
- Cloud Storage: Raw data
- Dataflow: Data Procesing
- AI Platform: Machine Learning
- App Engine: Cloud app hosting

3. Data Import and Storage

Import Dataset to CS and then BQ.
- Validations

4. Database Implementation

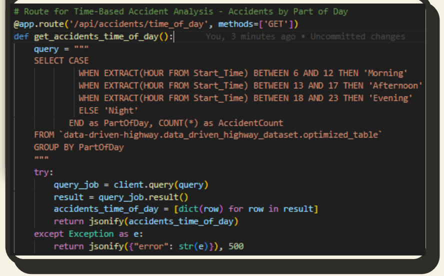
- We select data attributes based on key Metrics.
- Optimize squema for BigQuery

Cuadros Rivas, Alejandra - Cloud-Based Analytics Dashboard

Data Driven Highways Dashboard Blueprints

5. Data Processing and Analysis

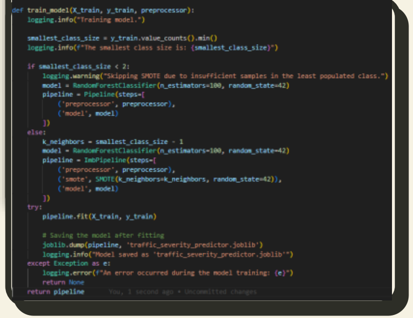
- Key Metrics: Accidents by Location
- API requests



```python
# Route for Time-Based Accident Analysis - Accidents by Part of Day
@app.route('/api/accidents/time_of_day', methods=['GET'])
def get_accidents_time_of_day():        You, 3 minutes ago • Uncommitted changes
    query = """
    SELECT CASE
             WHEN EXTRACT(HOUR FROM Start_Time) BETWEEN 6 AND 12 THEN 'Morning'
             WHEN EXTRACT(HOUR FROM Start_Time) BETWEEN 13 AND 17 THEN 'Afternoon'
             WHEN EXTRACT(HOUR FROM Start_Time) BETWEEN 18 AND 23 THEN 'Evening'
             ELSE 'Night'
           END as PartOfDay, COUNT(*) as AccidentCount
    FROM `data-driven-highway.data_driven_highway_dataset.optimized_table`
    GROUP BY PartOfDay
    """

    try:
        query_job = client.query(query)
        result = query_job.result()
        accidents_time_of_day = [dict(row) for row in result]
        return jsonify(accidents_time_of_day)
    except Exception as e:
        return jsonify({"error": str(e)}), 500
```

6. Machine Learning Implementation

- Predicts severity of accidents based on factors
- Random Forest Classifier
- Data loading, preprocess, model training, evaluation and saving

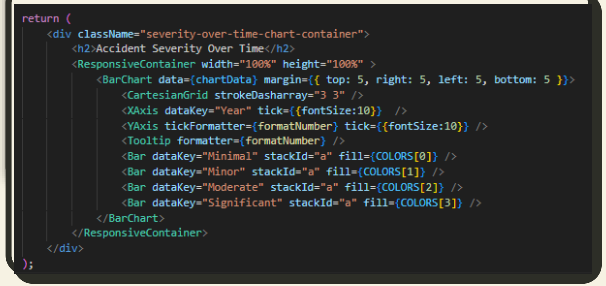


```python
def train_model(X_train, y_train, preprocessor):
    logging.info("Training model.")

    smallest_class_size = y_train.value_counts().min()
    logging.info(f"The smallest class size is: {smallest_class_size}")

    if smallest_class_size < 2:
        logging.warning("Skipping SMOTE due to insufficient samples in the least populated class.")
        model = RandomForestClassifier(n_estimators=100, random_state=42)
        pipeline = Pipeline(steps=[
            ('preprocessor', preprocessor),
            ('model', model)
        ])
    else:
        k_neighbors = smallest_class_size - 1
        model = RandomForestClassifier(n_estimators=100, random_state=42)
        pipeline = ImbPipeline(steps=[
            ('preprocessor', preprocessor),
            ('smote', SMOTE(k_neighbors=k_neighbors, random_state=42)),
            ('model', model)
        ])
    try:
        pipeline.fit(X_train, y_train)

        # Saving the model after fitting
        joblib.dump(pipeline, 'traffic_severity_predictor.joblib')
        logging.info("Model saved as 'traffic_severity_predictor.joblib'")
    except Exception as e:
        logging.error(f"An error occurred during the model training: {e}")
        return None
    return pipeline      You, 1 second ago • Uncommitted changes
```

7. UI/UX Design

- Core Functionalities
- Color Palette
- User-Friendly manner

8. Frontend Interface Development



```jsx
return (
    <div className="severity-over-time-chart-container">
        <h2>Accident Severity Over Time</h2>
        <ResponsiveContainer width="100%" height="100%" >
            <BarChart data={chartData} margin={{ top: 5, right: 5, left: 5, bottom: 5 }}>
                <CartesianGrid strokeDasharray="3 3" />
                <XAxis dataKey="Year" tick={{fontSize:10}}  />
                <YAxis tickFormatter={formatNumber} tick={{fontSize:10}} />
                <Tooltip formatter={formatNumber} />
                <Bar dataKey="Minimal" stackId="a" fill={COLORS[0]} />
                <Bar dataKey="Minor" stackId="a" fill={COLORS[1]} />
                <Bar dataKey="Moderate" stackId="a" fill={COLORS[2]} />
                <Bar dataKey="Significant" stackId="a" fill={COLORS[3]} />
            </BarChart>
        </ResponsiveContainer>
    </div>
);
```

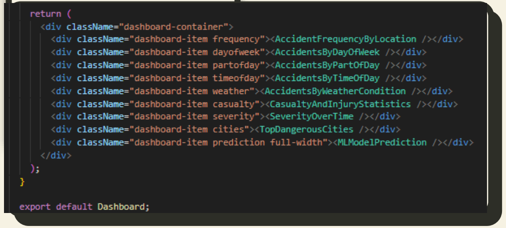
- React App
- Create a component for every route.

Cuadros Rivas, Alejandra - Cloud-Based Analytics Dashboard

Data Driven Highways Dashboard Blueprints

9. Backend and Frontend Integration

- JS and CSS files for the Dashboard 3x3 grid layout



```
    return (
      <div className="dashboard-container">
        <div className="dashboard-item frequency"><AccidentFrequencyByLocation /></div>
        <div className="dashboard-item dayofweek"><AccidentsByDayOfWeek /></div>
        <div className="dashboard-item partofday"><AccidentsByPartOfDay /></div>
        <div className="dashboard-item timeofday"><AccidentsByTimeOfDay /></div>
        <div className="dashboard-item weather"><AccidentsByWeatherCondition /></div>
        <div className="dashboard-item casualty"><CasualtyAndInjuryStatistics /></div>
        <div className="dashboard-item severity"><SeverityOverTime /></div>
        <div className="dashboard-item cities"><TopDangerousCities /></div>
        <div className="dashboard-item prediction full-width"><MLModelPrediction /></div>
      </div>
    );
  }

  export default Dashboard;
```

10. Testing

All the routes return the correct data, and the model correctly predicts based on the input it receives.

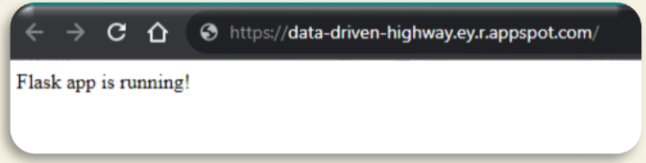
11. Deployment

- Proxy request failure
- inability of the frontend server (localhost:3000), to forward requests to the backend server at (localhost:5000).

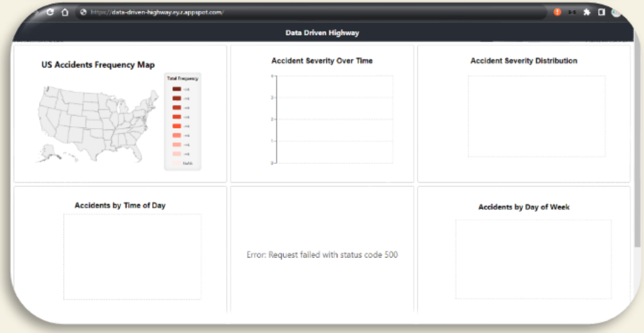
- Approach Cross Domain Request
  - VENV
  - Concurrently package
  - CORS: Cross origin Resource sharing
  - Docker

RESULTS

Faced Challenges in Deploying the System on Google Cloud

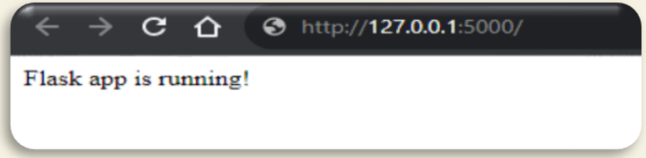


```
Flask app is running!
```

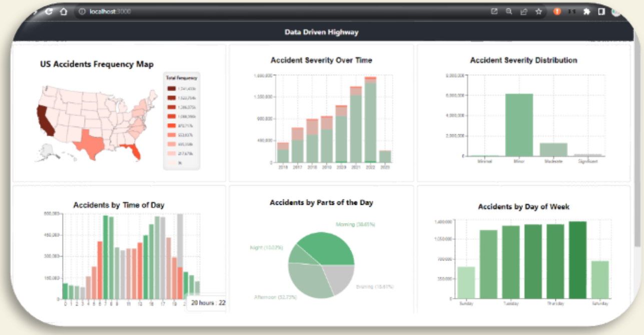

https://data-driven-highway.ey.r.appspot.com/
Data Driven Highway
US Accidents Frequency Map
Total Frequency
NaN
Accident Severity Over Time
4
3
2
1
0
Accident Severity Distribution
Accidents by Time of Day
Accidents by Day of Week

```
Error: Request failed with status code 500
```

Platform was successfully developed and local testing.

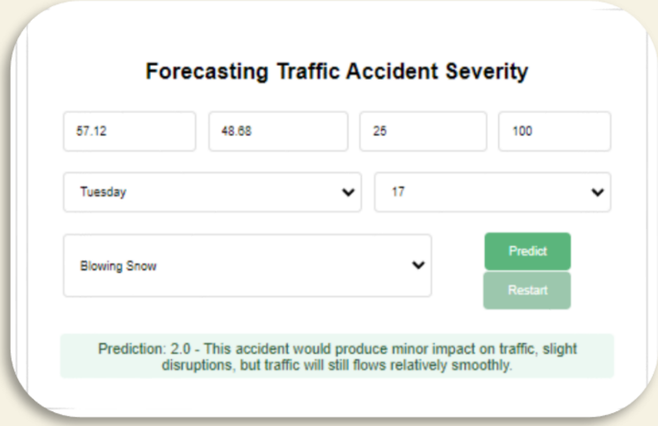


```
Flask app is running!
```



RESULTS

Enables data-driven decisión-making with predictive capabilities: Random Forest Classifier



Forecasting Traffic Accident Severity

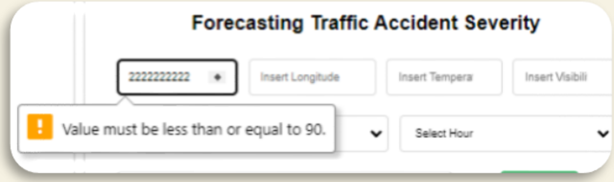
57.12 | 48.68 | 25 | 100

Tuesday ⌄ | 17 ⌄

Blowing Snow ⌄

Prediction: 2.0 - This accident would produce minor impact on traffic, slight disruptions, but traffic will still flows relatively smoothly.

Excel in performance and reliability during scalability and monitoring tests

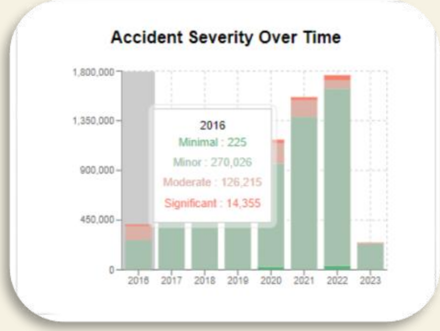


Forecasting Traffic Accident Severity

2222222222 | Insert Longitude | Insert Tempera | Insert Visibili

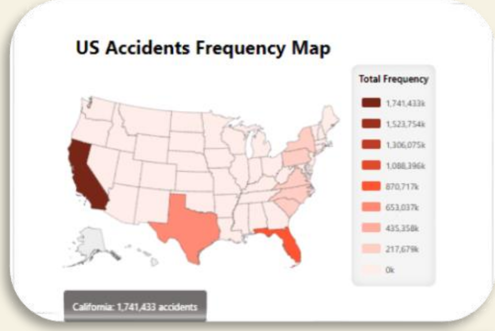
Value must be less than or equal to 90.

Select Hour ⌄



Accident Severity Over Time

2016
Minimal : 225
Minor : 270,026
Moderate : 126,215
Significant : 14,355



US Accidents Frequency Map

Total Frequency
1,741,433k
1,523,754k
1,306,075k
1,088,396k
870,717k
653,037k
435,358k
217,679k
0k

California: 1,741,433 accidents

Conclusions

Integrate various data sources onto a cloud platform for efficient real-time analysis.

Functional and advanced analytical dashboard, emphasizing data utility in traffic safety.

Utilized Python, Flask, React, and Google Cloud Big Query.

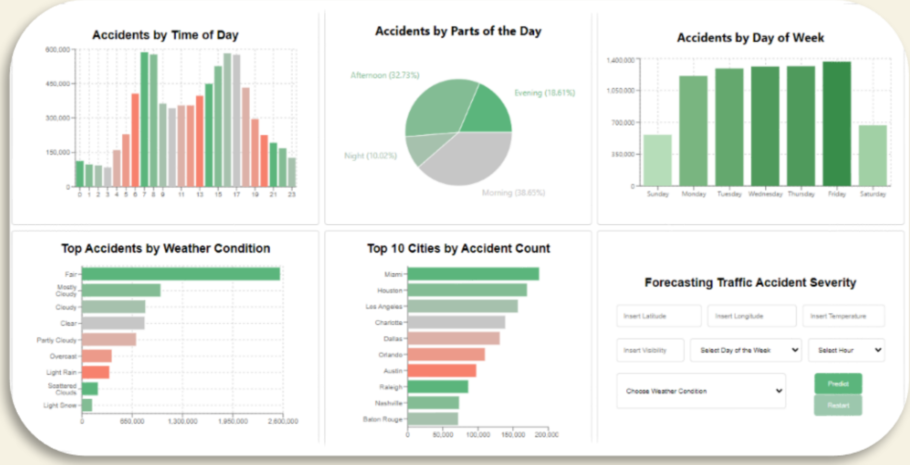
Integration of technologies despite deployment challenges.

Incorporated machine learning with a Random Forest Classifier for predictive insights.

Ongoing adaptability, problem-solving, and optimization in cloud computing and data analytics.

Future Development Opportunities

- Improve Cloud Deployment Processes
- Integrate real-time data with AI models
- Refine and adapt the user interface
- Regularly update for performance and scalability



Accidents by Time of Day

Accidents by Parts of the Day

Accidents by Day of Week

Top Accidents by Weather Condition

Top 10 Cities by Accident Count

Forecasting Traffic Accident Severity

Thank you for your attention!

Cuadros Rivas, Alejandra Paola

acuadrosrivas@edu.bme.hu