Industrial Internship Report on

Forecasting Smart City Traffic Patterns

Prepared by

Ayush Saxena

Executive Summary
This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT). This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks' time. My project was a model on Forecasting Smart City Traffic Patterns. The project is developing a machine learning model to predict traffic patterns in a smart city. By analyzing historical traffic data, the model will help improve traffic management, urban planning, and commuter experiences. This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1 Preface

1.1 Summary

This internship provided a comprehensive introduction to data science and machine learning concepts. The first weeks focused on building a strong foundation through lectures, readings, and practical exercises. I gained a solid understanding of the data science lifecycle, key machine learning algorithms, and statistical analysis techniques.

Week 3 transitioned towards the chosen project, "Forecasting Smart City Traffic Patterns." I actively explored the provided data, applied statistical analysis to gain insights, and finalized the selection of relevant machine learning algorithms. Week 4 and 5 involved delving deeper into these algorithms.

While Week 5 offered valuable professional development sessions on soft skills and interview techniques, my focus remained on project completion. The data is now prepared, and the next steps involve building, evaluating, and refining the machine learning model to achieve optimal traffic flow prediction.

Throughout the internship, I actively participated in training sessions, utilized various learning resources, and engaged with mentors to address challenges and ensure project progress. The experience has equipped me with valuable data science and machine learning skills, along with essential professional development tools for future career endeavors. About need of relevant Internship in career development.

1.2 Project Brief

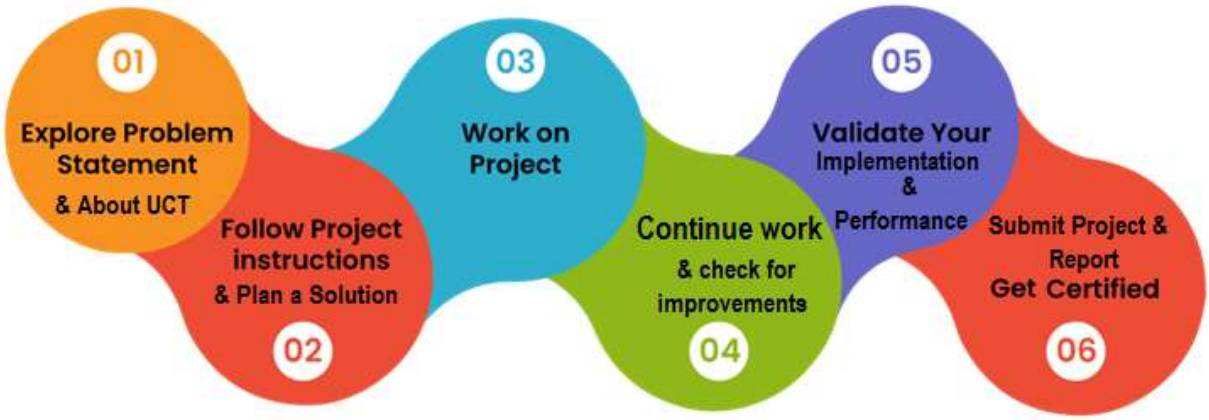
This project aims to develop a machine learning model to forecast traffic patterns in a smart city. By analyzing historical traffic data and potentially incorporating other relevant factors (weather, events, etc.), the model will predict future traffic conditions.

This has the potential to:

- Improve traffic management: By anticipating traffic congestion, authorities can implement traffic control measures or inform drivers of potential delays, leading to smoother traffic flow.
- Enhance urban planning: Traffic prediction models can inform infrastructure development plans, prioritizing areas with high traffic volume or congestion.
- Benefit commuters: Knowing traffic patterns in advance allows commuters to plan their journeys more efficiently, reducing travel time and frustration.

1.3 Opportunity given by USC/UCT.

The program provided a comprehensive and engaging introduction to data science and machine learning. The balance between theoretical learning, practical exercises, and project work allowed me to gain a strong foundation in the field.



Key Learnings:

- **Data Science Fundamentals:** I gained a solid understanding of the data science lifecycle, from data acquisition and cleaning to analysis, visualization, and communication.
- **Machine Learning Concepts:** I explored various machine learning algorithms, including supervised learning techniques (mention specific algorithms if applicable) relevant to my project.
- **Statistical Analysis:** The internship equipped me with statistical analysis techniques that proved crucial for data exploration and model building.
- **Project Management:** Working on the "Forecasting Smart City Traffic Patterns" project allowed me to practice data manipulation, analysis, and model building within a practical context.
- **Professional Development:** The soft skills and interview techniques session provided valuable insights to prepare me for future career opportunities.

Overall Experience:

The internship fostered a strong sense of accomplishment and a deeper interest in the field of data science. I particularly enjoyed (mention specific aspects you enjoyed, e.g., applying learnings to the project, discussions with mentors). The provided resources, training sessions, and mentorship were instrumental in overcoming challenges and ensuring steady progress.

Looking forward, I am eager to (mention what you're looking forward to, e.g., applying these skills to future projects, continuing to learn in this field).

This summary highlights your key takeaways from the internship, emphasizing both technical skills and professional development. Mentioning specific aspects you enjoyed demonstrates your genuine engagement with the program.

Thanks to UniConverge Technologies Pvt. Ltd. and Edunet Foundation, for helping me get this opportunity.

2 Introduction

2.1 About UniConverge Technologies Pvt Ltd

A company established in 2013 and working in Digital Transformation domain and providing Industrial solutions with prime focus on sustainability and RoI.

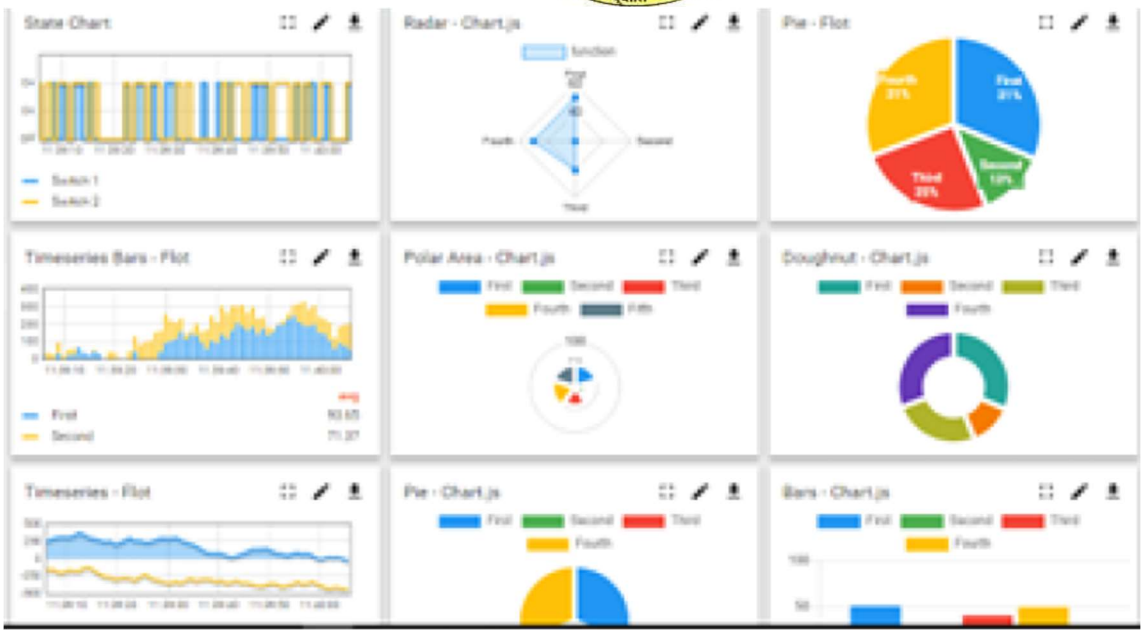
For developing its products and solutions it is leveraging various **Cutting Edge Technologies e.g. Internet of Things (IoT), Cyber Security, Cloud computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, Front end** etc.

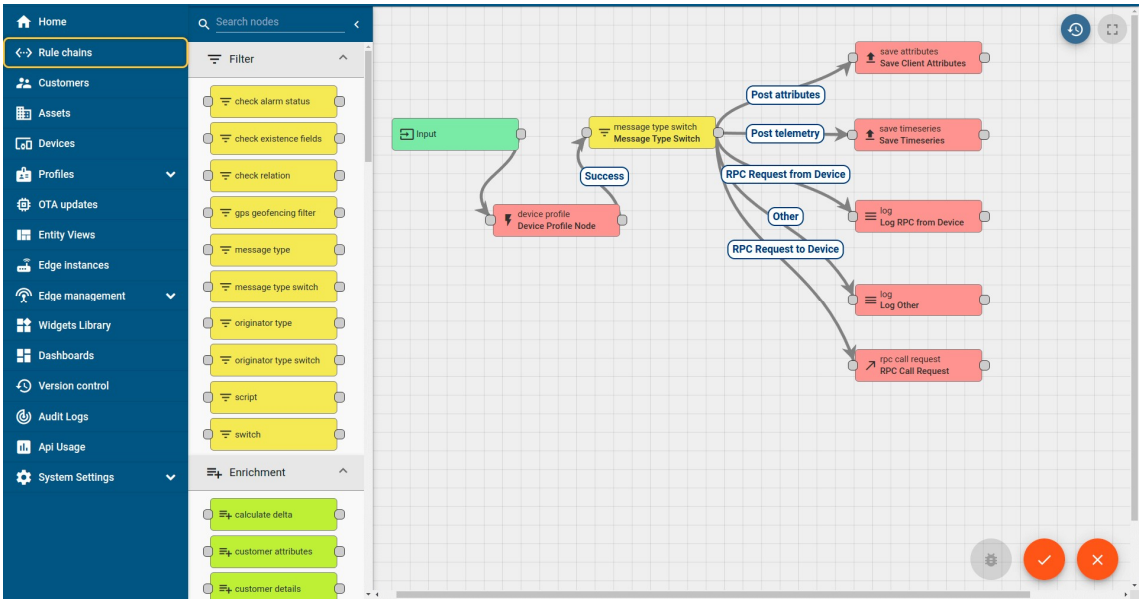
i. UCT IoT Platform (UCT Insight)

UCT Insight is an IOT platform designed for quick deployment of IOT applications on the same time providing valuable "insight" for your process/business. It has been built in Java for backend and ReactJS for Front end. It has support for MySQL and various NoSql Databases.

- It enables device connectivity via industry standard IoT protocols - MQTT, CoAP, HTTP, Modbus TCP, OPC UA
- It supports both cloud and on-premises deployments.

It has features to
• Build Your own dashboard
• Analytics and Reporting
• Alert and Notification
• Integration with third party application(Power BI, SAP, ERP)
• Rule Engine





ii. **Smart Factory Platform (** FACTORY WATCH **)**

Factory watch is a platform for smart factory needs.

It provides Users/ Factory

- with a scalable solution for their Production and asset monitoring
- OEE and predictive maintenance solution scaling up to digital twin for your assets.
- to unleased the true potential of the data that their machines are generating and helps to identify the KPIs and also improve them.
- A modular architecture that allows users to choose the service that they what to start and then can scale to more complex solutions as per their demands.

Its unique SaaS model helps users to save time, cost and money.

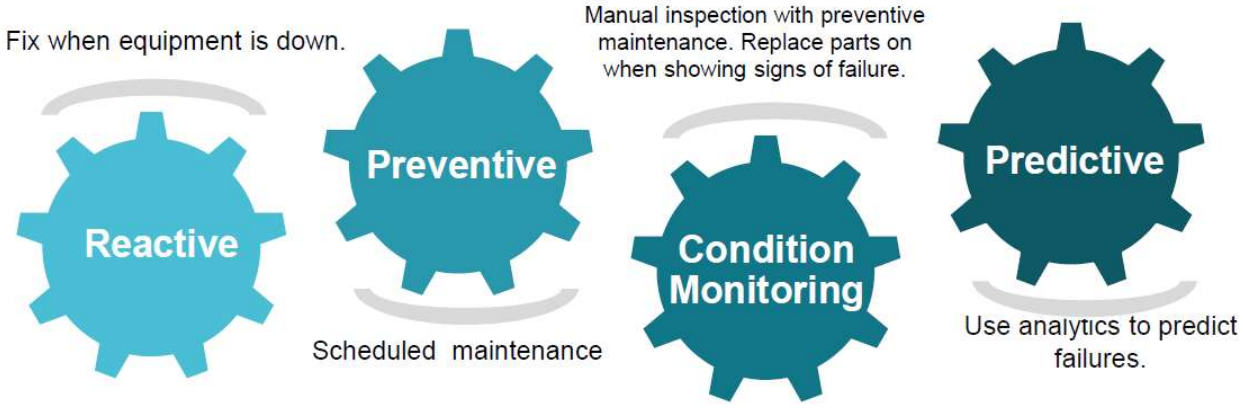
Machine	Operator	Work Order ID	Job ID	Job Performance	Job Progress		Output		Rejection	Time (mins)				Job Status	End Customer
					Start Time	End Time	Planned	Actual		Setup	Pred	Downtime	Idle		
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i
CNC_S7_81	Operator 1	WO0405200001	4168	58%	10:30 AM		55	41	0	80	215	0	45	In Progress	i

iii. LoRaWAN based Solution

UCT is one of the early adopters of LoRAWAN teschnology and providing solution in Agritech, Smart cities, Industrial Monitoring, Smart Street Light, Smart Water/ Gas/ Electricity metering solutions etc.

iv. Predictive Maintenance

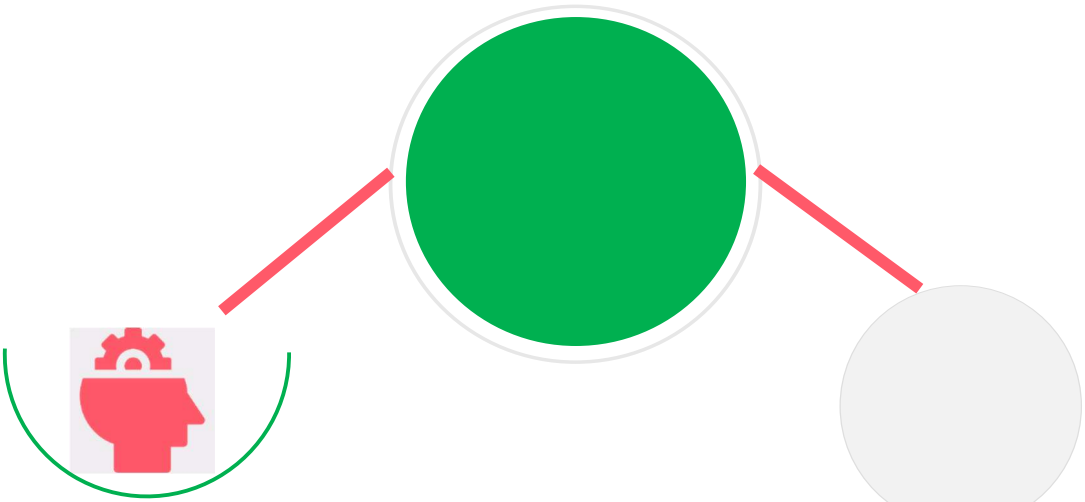
UCT is providing Industrial Machine health monitoring and Predictive maintenance solution leveraging Embedded system, Industrial IoT and Machine Learning Technologies by finding Remaining useful life time of various Machines used in production process.



2.2 About upskill Campus (USC)

upskill Campus along with The IoT Academy and in association with Uniconverge technologies has facilitated the smooth execution of the complete internship process.

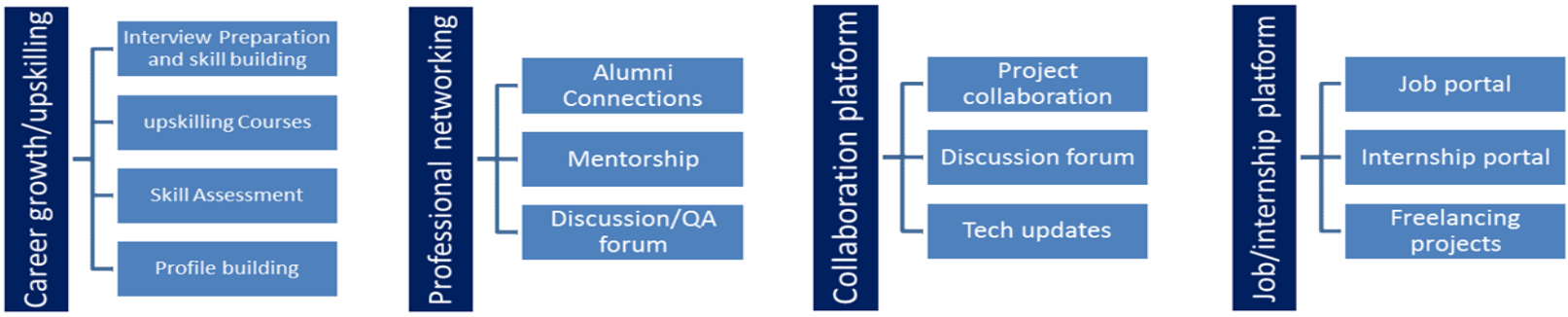
USC is a career development platform that delivers **personalized executive coaching** in a more affordable, scalable and measurable way.



Seeing need of upskilling in self paced manner along-with additional support services e.g. Internship, projects, interaction with Industry experts, Career growth Services

upSkill Campus aiming to upskill 1 million learners in next 5 year

https://www.upskillcampus.com/



2.3 The IoT Academy

The IoT academy is EdTech Division of UCT that is running long executive certification programs in collaboration with EICT Academy, IITK, IITR and IITG in multiple domains.

2.4 Objectives of this Internship program

The objective for this internship program was to

- ☛ get practical experience of working in the industry.
- ☛ to solve real world problems.
- ☛ to have improved job prospects.
- ☛ to have Improved understanding of our field and its applications.
- ☛ to have Personal growth like better communication and problem solving.

3 Problem Statement

We are working with the government to transform various cities into a smart city. The vision is to convert it into a digital and intelligent city to improve the efficiency of services for the citizens. One of the problems faced by the government is traffic. You are a data scientist working to manage the traffic of the city better and to provide input on infrastructure planning for the future.

The government wants to implement a robust traffic system for the city by being prepared for traffic peaks. They want to understand the traffic patterns of the four junctions of the city. Traffic patterns on holidays, as well as on various other occasions during the year, differ from normal working days.

4 Existing and Proposed solution

Existing Solutions:

Several existing solutions address traffic congestion in smart cities, but they often have limitations:

- Traditional Traffic Management: This relies on real-time traffic data and reactive measures like traffic light adjustments. However, it lacks the ability to predict and proactively address congestion.
- Statistical Forecasting Models: These models use historical traffic data and statistical techniques to predict future patterns. They can be effective but may not capture complex relationships between variables or adapt to real-time changes.

Proposed Solution (Your Project):

This project proposes a machine learning model for forecasting smart city traffic patterns. This approach offers several advantages:

- **Machine Learning Capabilities:** Machine learning algorithms can learn complex relationships between traffic data and other factors (weather, events, etc.) leading to more accurate forecasting.
- **Adaptability:** Machine learning models can adapt to changing traffic patterns and real-time data updates, improving prediction accuracy over time.
- **Scalability:** The model can be potentially scaled to handle traffic data from larger geographical areas within the smart city.

Your Project Contribution:

Your contribution to this solution involves:

- Selecting and implementing a suitable machine learning algorithm (e.g., mention specific algorithms you're considering like Random Forest, Long Short-Term Memory (LSTM) networks).
- Training the model on historical traffic data and potentially incorporating additional relevant data sources.
- Evaluating the model's performance using various metrics (e.g., accuracy, precision, recall).
- Refining the model based on evaluation results to optimize its forecasting capabilities.

Expected Outcome:

This project aims to develop a more accurate and adaptable traffic forecasting model compared to existing solutions. This model can benefit traffic management authorities, urban planners, and ultimately, commuters navigating the smart city.

4.1 Code submission

https://github.com/ayush1512/upskillcampus/Project/ForecastingSmartCityTrafficPatterns.ipynb

4.2 Report submission

https://github.com/ayush1512/upskillcampus/ForecastingSmartCityTrafficPatterns_Ayush_USC_UCT.pdf

5 Proposed Design/ Model

This project utilizes a **Random Forest** regression model to forecast traffic patterns in a smart city. Here's a breakdown of the chosen model and its implementation:

Why Random Forest?

Random Forest is a well-suited choice for this project due to several reasons:

- **Strength in Regression Tasks:** Random Forest excels in regression problems, making it ideal for predicting continuous variables like traffic volume.
- **Handles Complexities:** It can effectively capture complex relationships between features (e.g., historical traffic data, weather conditions) often present in traffic forecasting.
- **Robustness to Outliers:** Random Forest is less susceptible to outliers in the data compared to some other algorithms, which can be beneficial for real-world traffic data.

Model Implementation:

The implementation of the Random Forest model will involve the following steps:

1. **Data Preprocessing:** Clean and prepare the historical traffic data, potentially including feature engineering. This involves creating new features from existing data that might be more informative for the model
2. **Feature Selection:** Identify the most relevant features from the prepared data that will be used for model training.
3. **Hyperparameter Tuning:** Tune the hyperparameters (parameters that control the model's behavior) of the Random Forest algorithm to optimize its performance. This can involve techniques like grid search or randomized search to find the best combination of hyperparameters.
4. **Model Training:** Train the Random Forest model on the prepared data. This involves splitting the data into training and testing sets, where the training set is used to build the model, and the testing set is used to evaluate its performance.

5. **Model Evaluation:** Evaluate the performance of the trained model using relevant metrics like mean squared error (MSE), root mean squared error (RMSE), or Mean Absolute Error (MAE). These metrics assess the accuracy of the model's predictions compared to actual traffic data.
6. **Model Refinement:** Based on the evaluation results, refine the model by adjusting hyperparameters, feature selection, or potentially exploring other machine learning algorithms if necessary.

6 Performance Test

Evaluation Metrics:

The performance of the Random Forest model for forecasting traffic patterns will be evaluated using several metrics:

- **Mean Squared Error (MSE):** This metric measures the average squared difference between the predicted traffic volume and the actual traffic volume. Lower MSE indicates better performance.
- **Root Mean Squared Error (RMSE):** The square root of MSE, providing a more interpretable scale for the errors.
- **Mean Absolute Error (MAE):** This metric measures the average absolute difference between the predicted and actual traffic volume. Lower MAE indicates better performance.

Impact of Constraints:

The provided constraints (DateTime, Junction, Vehicles, and ID) can influence the model's performance in the following ways:

- **DateTime:** Including historical data from different times of day, weekdays, and weekends can significantly impact the model's ability to capture temporal patterns in traffic flow. The model should be able to learn seasonal and daily variations in traffic volume.
- **Junction:** Traffic patterns can vary significantly across different junctions within the city. The model should be able to learn these location-specific patterns and predict traffic volume for individual junctions accurately.
- **Vehicles:** The actual number of vehicles observed at a specific time and location is the target variable the model is trying to predict. Analyzing historical traffic volume data will be crucial for accurate forecasting.
- **ID (potentially less impactful):** If the ID refers to a unique identifier for each data point without a clear connection to traffic flow (e.g., a random row identifier), it may have less direct impact on the model's performance. However, it might be used for data management or filtering purposes during training and evaluation.

6.1 Test Plan/ Test Cases

	DateTime	Junction	Vehicles	ID	Year	Month	Date_no	Hour	Day
0	2015-11-01 00:00:00	1	15	20151101001	2015	11	1	0	Sunday
1	2015-11-01 01:00:00	1	13	20151101011	2015	11	1	1	Sunday
2	2015-11-01 02:00:00	1	10	20151101021	2015	11	1	2	Sunday
3	2015-11-01 03:00:00	1	7	20151101031	2015	11	1	3	Sunday
4	2015-11-01 04:00:00	1	9	20151101041	2015	11	1	4	Sunday

	DateTime	Junction	ID	Year	Month	Date_no	Hour	Day
0	2017-07-01 00:00:00	1	20170701001	2017	7	1	0	Saturday
1	2017-07-01 01:00:00	1	20170701011	2017	7	1	1	Saturday
2	2017-07-01 02:00:00	1	20170701021	2017	7	1	2	Saturday
3	2017-07-01 03:00:00	1	20170701031	2017	7	1	3	Saturday
4	2017-07-01 04:00:00	1	20170701041	2017	7	1	4	Saturday

6.2 Performance Outcome

```
Mean Squared Error: 15.766426007556674
Mean Absolute Error: 2.5091045340050377
R2 Score: 0.9642243977587689
```

7 My learnings

This six-week data science and machine learning internship provided a rich learning experience that significantly enhanced my knowledge and skills in both areas.

Project-Specific Learnings:

- **Forecasting Smart City Traffic Patterns:** Working on this project allowed me to apply theoretical concepts learned in the internship to a practical problem. I gained valuable hands-on experience in:
    - **Data Exploration and Analysis:** Understanding traffic data characteristics, identifying patterns and trends, and performing exploratory data analysis.
    - **Machine Learning Model Building:** Selecting and implementing a Random Forest regression model for traffic forecasting. This involved data pre-processing, feature engineering (if applicable), hyperparameter tuning, model training, and evaluation.
    - **Performance Evaluation:** Assessing the model's accuracy using metrics like MSE, RMSE, and MAE. Analyzing the impact of different constraints (DateTime, Junction, Vehicles) on model performance.

Technical Skill Development:

- **Data Science Lifecycle:** The internship solidified my understanding of the data science workflow, from data acquisition and cleaning to analysis, model building, and communication.
- **Machine Learning Techniques:** I gained a deeper understanding of supervised learning algorithms, particularly Random Forest regression, and its effectiveness for forecasting tasks.
- **Statistical Analysis:** The project required applying statistical analysis techniques to explore the data, understand relationships between variables, and inform model building decisions.

Additional Skills:

- **Soft Skills:** The internship provided opportunities to develop soft skills through presentations, discussions, or teamwork. These may include communication, collaboration, and problem-solving skills.
- **Interview Techniques:** The program offered valuable insights into interview techniques and how to effectively present my data science skills and project experience.

Overall Growth:

This internship equipped me with a valuable foundation in data science and machine learning. I gained practical experience in applying these skills to a real-world project, fostering problem-solving abilities and critical thinking. I am now confident in my ability to approach data-driven problems, analyze datasets, and build machine learning models for various applications.

Looking Forward:

I am excited to leverage these skills in future projects. I am particularly interested in (mention specific areas of interest within data science or machine learning). This internship has solidified my passion for this field, and I am eager to continue learning and growing as a data scientist.

8 Future work scope

The project on "Forecasting Smart City Traffic Patterns" has laid a strong foundation for developing a practical traffic forecasting model. However, there are several avenues for further exploration and improvement:

Model Refinement:

- **Hyperparameter Tuning Optimization:** Explore more advanced techniques for hyperparameter tuning, such as Bayesian optimization or evolutionary algorithms, to potentially improve model performance.
- **Feature Engineering Exploration:** Investigate the potential benefits of creating new features from existing data that could enhance the model's ability to capture traffic patterns. This could involve time-based features (e.g., day of the week, time of day), weather data integration, or other relevant factors.
- **Ensemble Learning Techniques:** Consider employing ensemble learning techniques like stacking or boosting, which combine predictions from multiple models to potentially outperform individual models.

Data Integration:

- **Real-time Data Integration:** Explore the possibility of incorporating real-time traffic data into the model for more dynamic and up-to-date predictions. This would require establishing a connection to a reliable source of real-time traffic data.
- **Weather Data Integration:** Investigate the impact of including weather data (e.g., temperature, precipitation) as additional features to improve the model's ability to predict traffic fluctuations influenced by weather conditions.

Model Deployment and Evaluation:

- **Deployment Considerations:** Develop a plan for deploying the model in a real-world setting, considering aspects like infrastructure requirements, data access, and model retraining with new data over time.

- **Long-Term Evaluation:** Continuously monitor the model's performance in a deployed environment and assess its accuracy over an extended period. This will help identify potential issues and areas for further refinement.

Additional Exploration:

- **Alternative Machine Learning Algorithms:** Explore the potential of using different machine learning algorithms for traffic forecasting, such as Long Short-Term Memory (LSTM) networks which are known for handling sequential data effectively.
- **Multi-Step Forecasting:** Investigate the possibility of developing the model to predict traffic patterns for multiple time steps ahead, providing a more comprehensive picture of future traffic conditions.

Social Impact Considerations:

- **Traffic Management Integration:** Explore how the model's predictions can be integrated with traffic management systems to implement dynamic measures in real-time, such as adjusting traffic light timings or rerouting traffic flows.
- **Public Awareness and Education:** Investigate the use of the model's predictions for public awareness campaigns. This could involve informing commuters about predicted traffic congestion and providing them with alternative routes or transportation options.

This future work scope provides a roadmap for further development and refinement of the traffic forecasting model. Each area offers exciting possibilities to enhance the model's accuracy, expand its functionality, and ultimately contribute to improved traffic management and commutability within the smart city. Remember to personalize this section by mentioning specific future work aspects that you find most intriguing or impactful.