AI Berth Sharing: Elevating Indian Railways Journeys: Prototype

Satyam Sharma

satyamdelhi120@gmail.com

Intern (Solo) at

FeyNN labs

ABSTRACT-

The report delineates an initiative aimed at mitigating overcrowding within unreserved (U/R) compartments of the Indian Railways through the application of machine learning (ML). It encompasses several key components, including prototype selection, development, business modeling, and financial modeling. Commencing with the selection of ML algorithms to identify passengers amenable to sharing berths, the initiative progresses to the development of prototypes with a strong emphasis on user-friendly interfaces tailored to both passengers and ticket collectors. The overarching business model is designed to elevate the travel experience of middle-class passengers by offering reduced reservation charges and optimizing seat allocation to bolster revenue generation. Furthermore, the financial analysis conducted assesses the impact on revenue, ticket sales, fines reduction, and operational costs, thereby evaluating the feasibility and return on investment of the initiative. By leveraging ML capabilities, the initiative endeavours to alleviate overcrowding, enhance passenger experience, and generate value, while concurrently presenting avenues for the implementation of efficient AI-driven operations..

I) Prototype Selection-

Motive behind the idea-

Everyone enjoys traveling by train, but only when you have your reserved seat in hand. However, consider scenarios where you have to travel a distance of 4 to 5 hours and, at the last moment, are unable to reserve a seat. Unfortunately, you must travel in the general coach, crossing your fingers for just a piece of seat to sit on. This is a significant problem for many people because the majority in India prefer train travel due to several reasons: A) it is less expensive, B) it is safe (especially for women), and C) it offers good connectivity. To make the journey more comfortable, we should introduce an AI-driven solution that helps secure seats and also provides strong incentives for other passengers to share their berth due to reduced ticket prices.

Problem statement-

The unreserved (U/R) compartments within Indian Railways are plagued by significant overcrowding, posing considerable challenges for ticket collectors in conducting efficient ticket checks. This overcrowding not only detracts from the overall travel experience for passengers but

also complicates the enforcement and imposition of fines for unauthorized travel. Additionally, the inability of middle-class passengers to secure sleeper or 3A reservations creates a notable demand-supply gap, thereby impacting passenger comfort and hindering revenue generation potential for Indian Railways. Addressing this pressing issue calls for an innovative solution that harnesses technology to optimize seat utilization, streamline ticket checks, and enhance revenue streams for both Indian Railways and the implementing company.

1. Feasibility

Numerous companies have established partnerships with Indian Railways for ticketing, catering, and other services. Several critical steps must be addressed to ensure the feasibility of this proposed model:

a. Data collection from Indian Railways-
   Historically, no operator or company has actively pursued this demand from Indian Railways. As a company, it is imperative to assure Indian Railways of the secure usage of passenger data for this service. Building trust and ensuring data privacy will be pivotal in establishing a successful partnership. Additionally, it is crucial to note that the timeframe required to initiate this business idea is limited to the duration of *tying up with Indian Railways* and *collecting the necessary data to build a sustainable AI model*. This process should ideally be completed within a timeframe of ***two years***.

b. This addition emphasizes the importance of a prompt initiation of the business idea after securing the necessary partnership with Indian Railways and underscores the expected timeframe for data collection and model development. Passengers' Willingness to Share Their Berths:

Another crucial factor determining the feasibility of the proposal hinges on passengers' willingness to share their berths. The success of this model is contingent upon passengers' readiness to participate. Understanding and analyzing passenger attitudes, particularly in a country like India with diverse demographics and income levels, are paramount. Despite potential challenges, this model holds promise for sustainability with careful consideration of passenger preferences and needs.

2. Viability-

Market demand: There's a clear market demand indicated by the existing overcrowding issue and the desire for improved travel experiences among passengers. This indicates a potential willingness to adopt and pay for a solution that addresses these pain points.

Revenue Generation: We recognize the importance of generating additional revenue streams for both Indian Railways and our company. By implementing premium services and optimizing seat allocation, we aim to maximize revenue potential while simultaneously improving passenger satisfaction and loyalty.

Regulatory Compliance: Compliance with all legal and regulatory requirements is paramount to the success of our proposal. We are committed to following best practices and collaborating closely

with regulatory authorities to ensure full compliance, thereby mitigating risks and enhancing the long-term viability of our solution.

Competitive Landscape: In a competitive market environment, differentiation is key to success. Our unique value proposition, coupled with our commitment to innovation and customer satisfaction, sets us apart from competitors and positions us as a leader in addressing the challenges of overcrowding in Indian Railways. Additionally, it is essential to note that this *business idea has the potential to yield significant benefits for a sustained period*, ***spanning at least 6 to 7 years in its initial state***. While competition is a factor to consider, our proactive approach includes continuously conditioning our model to align with evolving market demands, ensuring our continued relevance and dominance in the industry.

Constraints-

1. Night Travel Hindrance (Primary Hurdle): The likelihood of berth sharing diminishes significantly as all travel occurs during ***nighttime***. Passengers may be reluctant to share their berths in these conditions.
2. Consistent Data Management Requirement: Due to the integration of machine learning, our model relies on consistent and efficient data management. Continuous updates and maintenance are essential for optimal functionality.
3. Extended Implementation Timeline: Achieving full implementation is projected to span the next 2-4 years. Given the *novelty* of this business model, the absence of precedence necessitates the collection of new data features, such as the 'willingness of passengers to share berths,' for an enhanced recommendation model.
4. Complete Reliance on Willing Middle-Class Customers: The success of our model is contingent upon the willingness of middle-class customers to pay a slightly higher fare for a more comfortable journey. If this willingness is lacking, it directly impacts the company's revenue.

3. Monetization-

Our proposed monetization strategy offers a scalable approach to generating revenue while simultaneously addressing the issue of overcrowding in Indian Railways' unreserved compartments. By charging passengers for the option to upgrade to upper-level coaches such as sleeper or AC Tier 3, we not only create a new revenue stream but also incentivize passengers to opt for a more comfortable travel experience.

95.3% travellers opted for general and sleeper classes

Breaking down the passenger distribution, the numbers mirror a consistent trend, with 95.3 percent opting for general and sleeper classes, while 4.7 percent preferred AC classes. “This shows that 95.3 percent of passengers, out of the total number between April and October 2023, travelled in general and sleeper class and 4.7 passengers chose AC classes,” a railway official said. A railway official highlighted this trend, emphasising that the majority of passengers between April and October 2023 favored general and sleeper class accommodations.

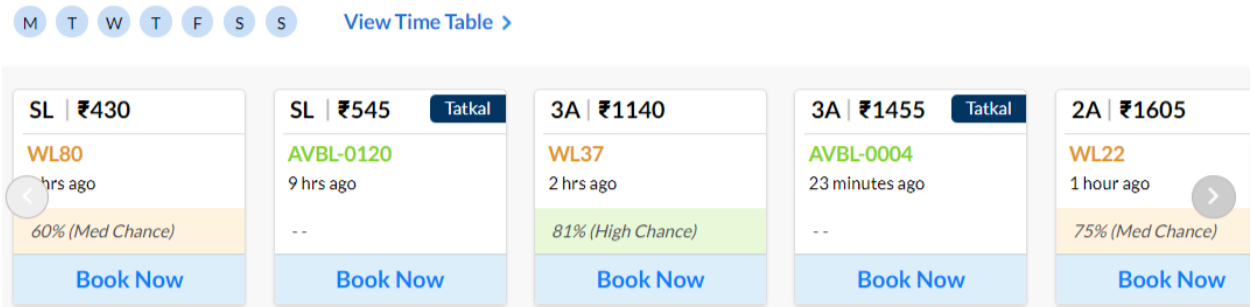
Target Market: Our target market comprises passengers who belong to the upper echelons of the middle class and prefer traveling in non-AC coaches.

Revenue Potential:

Let’s consider an example to illustrate the revenue potential of our monetization strategy.

Passenger Mahesh, a junior data scientist earning between 50,000 to 60,000 rupees per month, wishes to travel from Varanasi to Delhi by train due to poor connectivity and cost-effectiveness. Despite being unable to reserve a seat, Mahesh opts for an upgrade to an upper-level coach. Satyam, who has already reserved a seat in AC3, agrees to share his berth with Mahesh during the daytime journey. The additional amount Mahesh pays for the upgrade is calculated as two-thirds of the price difference between sleeper and AC3, amounting to approximately 473 rupees. Out of this amount, the railways receive 20%, which is around 94 rupees. Satyam receives 60% of the amount, approximately 282 rupees, while our company also receives 20%, totaling 94 rupees.

This monetization strategy not only benefits passengers like Mahesh but also provides additional revenue streams for Indian Railways and our company.



Net amount received by *Indian Railways, Service provider, Satyam.*

	Indian Railways	Service Provider	Mahesh	Satyam
Debit	Nil	Nil	473	Nil
Credit	94	94	Nil	282
Net	+ 94	+94	430+473=-903	1140-282=-858

The primary reason for ensuring that Satyam's total cost remains lower than Mahesh's is rooted in the nature of this business model, where the foremost consideration is Satyam's willingness to share his berth, rather than Mahesh's need for a seat. To render this arrangement viable, the critical factor lies in enticing Satyam by providing a reduced fare.

Scalability:

The scalability of our monetization strategy lies in its capacity to cater to a broad spectrum of passengers seeking a more comfortable travel experience. As an increasing number of passengers opt for the upgrade option, our revenue potential scales proportionally. Moreover, through the adept utilization of technology and efficient pricing models, we can seamlessly expand our operations to meet burgeoning demand across various routes and regions.

Our monetization strategy not only benefits passengers by providing them with a superior travel option but also fosters a mutually beneficial scenario for Indian Railways and our company. Passengers enjoy enhanced travel comfort, Indian Railways garners supplementary revenue without substantial investment, and our company earns a commission for facilitating the upgrade process.

II) **Business Modelling-**

Customer assessment:

Our potential customers are individuals in India, particularly those belonging to the middle class, who are seeking a more comfortable and convenient travel experience within the constraints of their average monthly income, which typically ranges from 10,000 to 12,000 INR. Given the current living standards in India (ranked at 56 in quality of life), where cost considerations play a crucial role, train travel is a preferred and economical mode of transportation.

The crowded conditions in General/UR coaches of Indian Railways, coupled with the limited availability of these coaches, have created challenges for passengers seeking a more pleasant journey. Many individuals are willing to pay a premium for enhanced comfort and convenience, even if it exceeds the minimal cost of a general ticket. This presents an opportunity for our service to cater to those passengers who value a more relaxed travel experience and are willing to contribute to the optimization of seat utilization, thereby increasing overall revenue.

Market assessment:

In the current market landscape, there are several authorized partners of IRCTC, such as Paytm, Ixigo, Moovit, Trainline, and RailYatri. These partners have successfully provided seamless experience for train bookings, including tatkal ticket reservations. Additionally, IRCTC has expanded its services through authorized partners like Zomato, enhancing food choices for railway passengers.

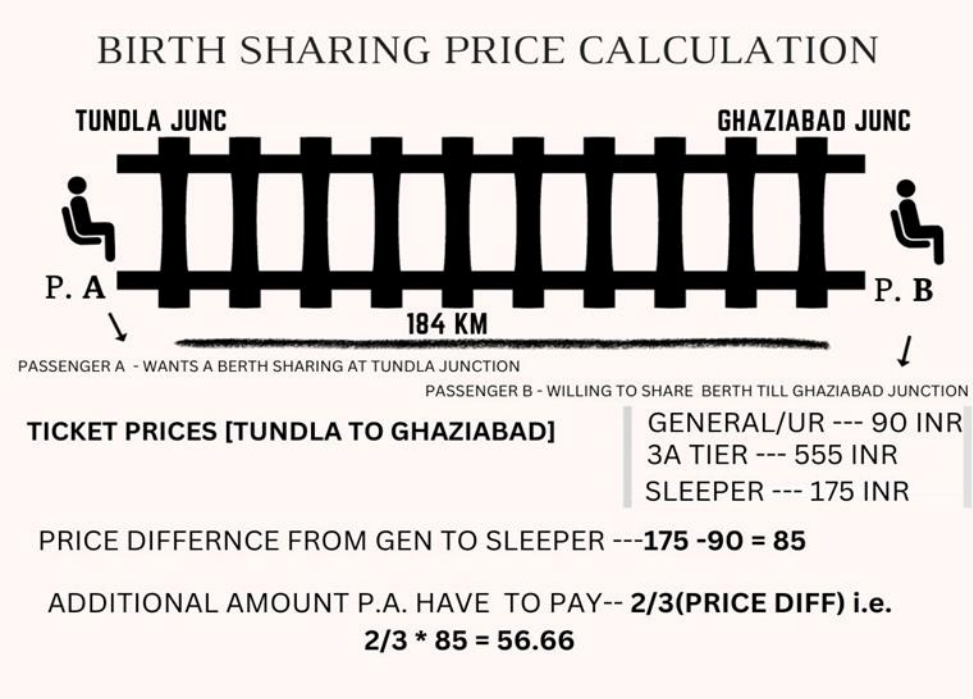
While Paytm leads with a 14% market share in IRCTC ticket booking, the proposed business idea remains unexplored by existing authorized partners. This presents a unique opportunity for any new player entering the market with a ticket booking option integrated with the innovative concept. Given the apparent gap in the market, there is potential to capture a significant share, tapping into the unmet needs of passengers seeking improved comfort and convenience during train travel. The absence of similar services from existing partners indicates a potential market demand for the proposed solution.

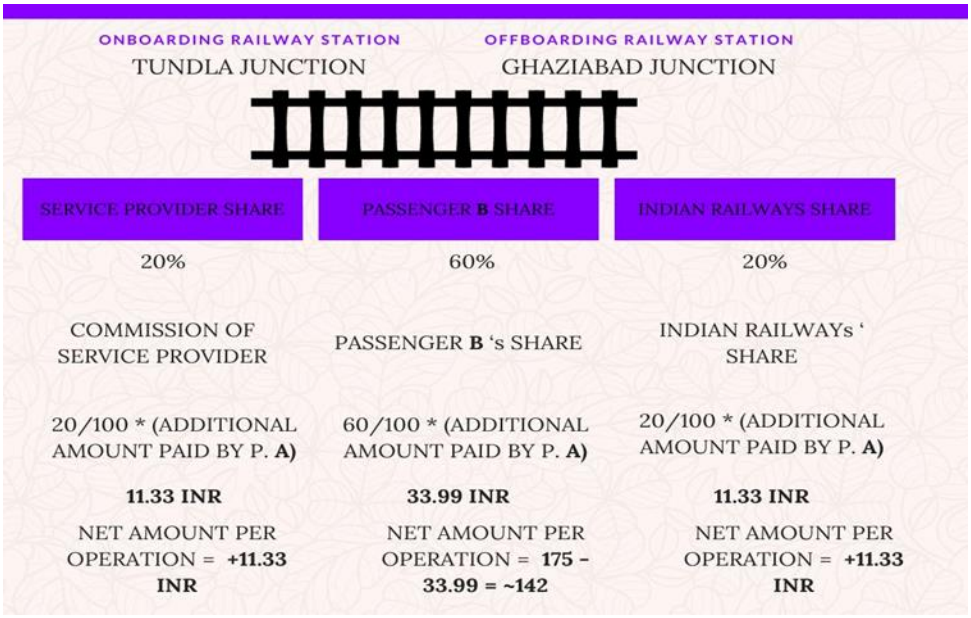
In our business model, the financial transactions involve three key entities:

- Indian Railways
- Passengers Utilizing the Service
- Service Provider (Our Company)

Our company will charge commission solely for the services provided, forming the basis of the financial transactions within this model.

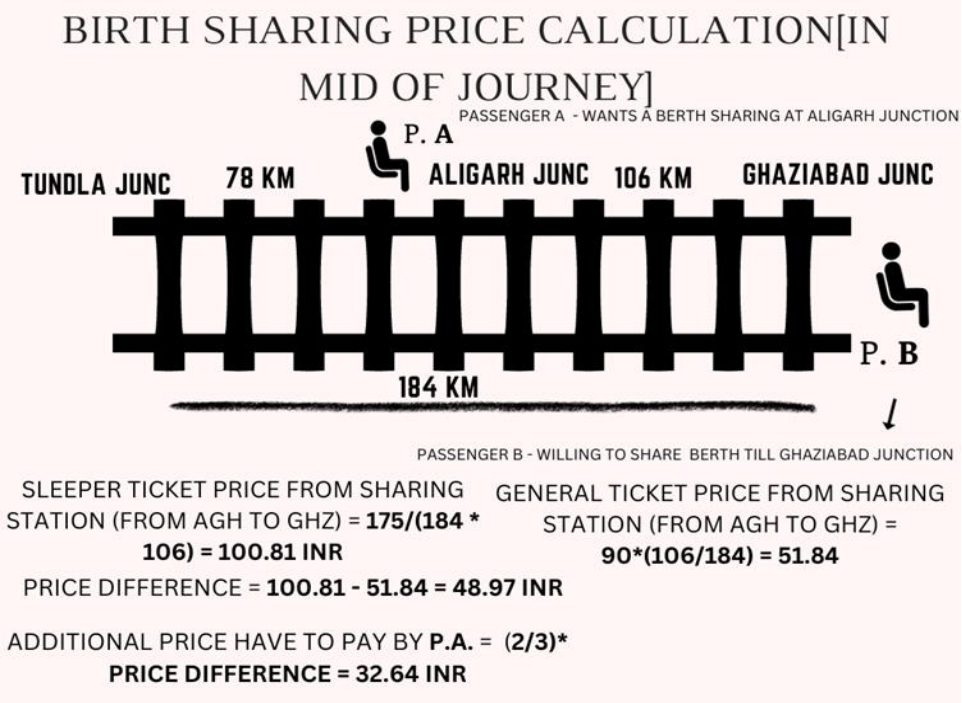
1. Monetization-

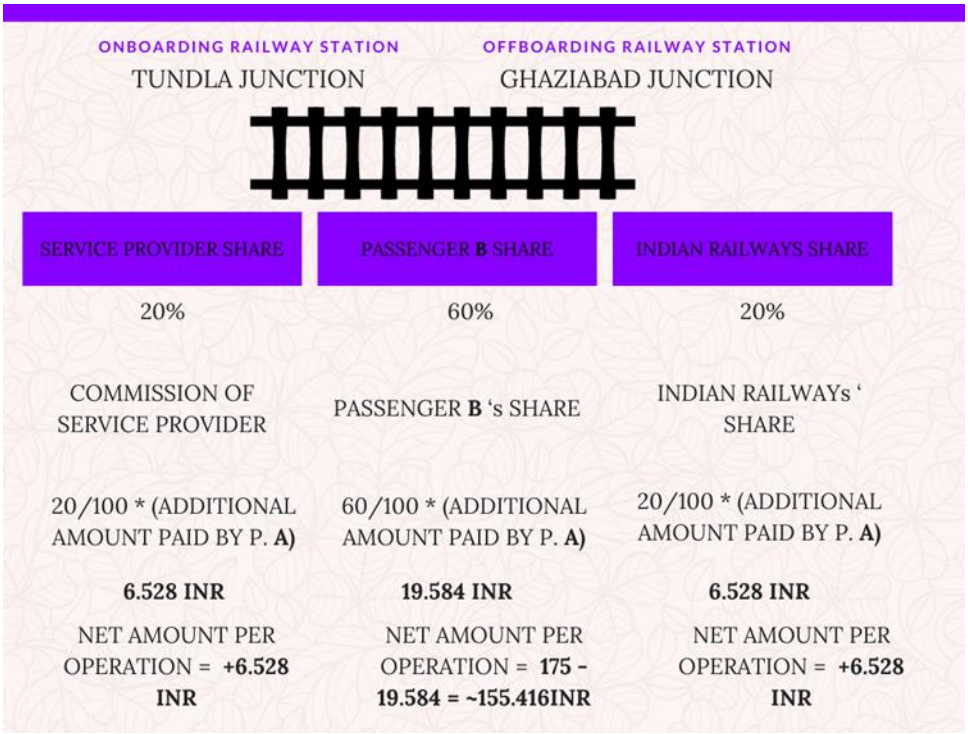




Entities	PAID (-)[INR]	RECEIVE (+)[INR]	NET [INR]
Passenger A	(-90) + (-56.66)	0	-146.66
Passenger B	-175	+33.99	~-141.01
Service Provider	0	+11.33	+11.33
Indian Railways	0	+11.33	+11.33

Upon examining the calculations provided above, it is essential to consider scenarios where Passenger A seeks berth sharing not from the origin of their journey but from an **intermediate station**. In such cases, the calculation will be as follows –





Entities	PAID (-)[INR]	RECEIVE (+)[INR]	NET [INR]
Passenger A	(-90) + (-32.64)	0	-122.64

Passenger B	-175	+19.584	-155.416
Service Provider	0	+6.528	+6.528
Indian Railways	0	+6.528	+6.528

III) Prototype Development

Note: ***As we lacked the required dataset necessary to successfully realize this business idea, we encountered a challenge in prototype development. Consequently, we resorted to forcefully injecting data columns into the dataset, restricting our model's ability to derive meaningful insights from all available features. This limitation led to the development of a simplistic model reliant solely on one independent feature. It is essential to note that the primary objective of this prototype is to provide a conceptual overview of how we intend to embark on this business journey, rather than to procedure a fully functional code module.***

We will utilize the Python programming language for implementing machine learning models.

Dataset Overview:

```
df.head(5)
```

	Train No	Train Name	SEQ	Station Code	Station Name	Arrival time	Departure Time	Distance	Source Station	Source Station Name	Destination Station	Destination Station Name
0	107	SWV-MAO-VLNK	1	SWV	SAWANTWADI R	00:00:00	10:25:00	0	SWV	SAWANTWADI ROAD	MAO	MADGOAN JN.
1	107	SWV-MAO-VLNK	2	THVM	THIVIM	11:06:00	11:08:00	32	SWV	SAWANTWADI ROAD	MAO	MADGOAN JN.
2	107	SWV-MAO-VLNK	3	KRMI	KARMALI	11:28:00	11:30:00	49	SWV	SAWANTWADI ROAD	MAO	MADGOAN JN.
3	107	SWV-MAO-VLNK	4	MAO	MADGOAN JN.	12:10:00	00:00:00	78	SWV	SAWANTWADI ROAD	MAO	MADGOAN JN.
4	108	VLNK-MAO-SWV	1	MAO	MADGOAN JN.	00:00:00	20:30:00	0	MAO	MADGOAN JN.	SWV	SAWANTWADI ROAD

- The dataset encompasses comprehensive information regarding the entire travel journey from the source station to the destination station.
- It includes the arrival and departure times of trains at all stations where the train makes a stop.
- Distances between consecutive stations are also provided.
- The dataset further includes the name of the destination station.
- Each entry is identified by a serial number (SEQ), denoting the sequence of stations traversed by the train.

- For a specific sequence of SEQ, corresponding train numbers and names are provided, delineating the train operating on that route.

  Note: Our dataset requires extensive preprocessing and filtering before it can be utilized to develop a machine learning model. The following steps outline the preprocessing tasks we will undertake.

- Generate a new feature representing the duration of the journey.
- Introduce a new feature named 'night_travelling' to distinguish between journeys occurring during nighttime and daytime.

Data preprocessing-

```
# checking null values
df.isna().sum()
```

```
Train No                      0
Train Name                    0
SEQ                           0
Station Code                  0
Station Name                  0
Arrival time                  5
Departure Time                5
Distance                      5
Destination Station          10
Destination Station Name     10
dtype: int64
```

Drop Null values

```
df.dropna(inplace=True)
```

```
df.isna().sum()
```

```
Train No                     0
Train Name                   0
SEQ                          0
Station Code                 0
Station Name                 0
Arrival time                 0
Departure Time               0
Distance                     0
Destination Station          0
Destination Station Name     0
dtype: int64
```

- There are a couple of key steps undertaken in the preprocessing function to prepare the dataset for machine learning analysis:
- Time Format Conversion: Initially, the *Arrival time* and *Departure Time* columns are transformed into the actual time format using the pd.to_datetime() function.
- Identification of Intermediate Stations: The dataset is segmented to exclude the initial and final stations, thereby isolating the intermediate stations along the train route.

- Creation of *night_travelling* Feature: A novel feature, *night_travelled,* is engineered to denote whether the journey transpired during nighttime or daytime based on the arrival time.
- Aggregation and Grouping: The intermediate station data is aggregated and grouped by train number, consolidating pertinent information such as train name, sequence, station name, arrival and departure times, distance, and destination station.
- Data Refinement: Finally, extraneous columns like *Station Code* are removed, and the index is reset to facilitate further analysis and modeling. The refined dataset, optimized for machine learning tasks, is then returned for subsequent processing.

```
def preprocessing(df):
  # Converting into actual time format
  df['Arrival time'] = pd.to_datetime(df['Arrival time'])
  df['Departure Time'] = pd.to_datetime(df['Departure Time'])

  intermediate_stations = df.groupby('Train No').apply(lambda x: x.iloc[1:-1])

  # creating new feature 'Night travelling' 1 for yes 0 for No
  intermediate_stations['night_travelled'] = intermediate_stations['Arrival time'].apply(lambda x: 1 if x.hour >= 21 or x.hour < 6 else 0)

  # Group by index and aggregate the columns
  grouped_df = intermediate_stations.groupby(level=0).agg({
    'Train Name': 'first',
    'SEQ': 'first',
    'Station Code': 'first',
    'Station Name': 'first',
    'Arrival time': 'first',
    'Departure Time': 'last',
    'Distance': 'sum',
    'Destination Station': 'last',
    'Destination Station Name': 'last',
    'night_travelled': 'max'
})

  # Reset the index to make 'Train No' a column again
  grouped_df.reset_index(inplace=True)

  # Remove the 'Station Code' column
  grouped_df.drop(columns=['Station Code'], inplace=True)

  return grouped_df
```

- Grouping the dataset by trains was necessary due to the excessive number of rows associated with each train, thereby streamlining the dataset for more efficient analysis.

- The 'Station Code' feature was dropped as it was deemed unnecessary for predictive modeling purposes. This decision was made to focus solely on independent features crucial for predicting outcomes, thereby refining the dataset to include only pertinent information.

```
[ ] processed_df['night_travelled'].value_counts()

    0    5409
    1    4456
    Name: night_travelled, dtype: int64
```

- The data is now fully balanced, comprising 5409 instances with a value of 0 (indicating no night travelling) and 4456 instances with a value of 1 (indicating night travelling).

Data insertion-

It is a pivotal step in this process, as it involves acquiring and integrating the dataset essential for this business idea.[Refer this article – to know more about feature selection]

```
[11] # Since we do not have complete dataset so we have to make it
     processed_df['Willing_to_share'] = processed_df['night_travelled'].apply(lambda x : 1 if x ==1 else 0)
```

```
processed_df.head(5)
```

	Train No	Train Name	SEQ	Station Name	Arrival time	Departure Time	Distance	Destination Station	Destination Station Name	night_travelled	Willing_to_share
0	22989	BDTS MHV WEE	10	BOTAD	2024-02-28 23:14:00	2024-02-28 01:55:00	4621	MHV	MAHUVA	1	1
1	22990	MHV BDTS SF	2	RAJULA ROAD	2024-02-28 13:27:00	2024-02-28 03:57:00	4851	BDTS	BANDRA TERMINUS	1	1
2	22991	BDTS VRL SF	2	BORIVLI	2024-02-28 13:11:00	2024-02-28 03:00:00	4938	VRL	VERAVAL	1	1
3	22992	VRL BDTS SF	2	JUNAGADH JN	2024-02-28 14:15:00	2024-02-28 04:35:00	4161	BDTS	BANDRA TERMINUS	1	1
4	22993	BDTS MHV SF	2	BORIVLI	2024-02-28 15:56:00	2024-02-28 06:24:00	6692	MHV	MAHUVA	1	1

- Here, we have recorded our new feature values by utilizing the condition of the 'night_travelled' feature.

Note: This dataset is not based on real-world data; rather, values have been generated using a fixed condition. Consequently, our model will rely solely on one feature, namely 'night_travelled', for predicting outcomes. However, in practical scenarios, all independent features should play a significant role in predicting the outcome value.

Data encoding-

```
[89] from sklearn.model_selection import train_test_split
     X_train,X_test, y_train,y_test = train_test_split(X,y, random_state=40)

     # shape and size
     print('Input training dataset shape {}'.format(X_train.shape))
     print('Input test dataset shape {}'.format(X_test.shape))
     print('Output training dataset shape {}'.format(y_train.shape))
     print('Output test dataset shape {}'.format(y_test.shape))
```

```
Input training dataset shape (2767, 4)
Input test dataset shape (923, 4)
Output training dataset shape (2767, 1)
Output test dataset shape (923, 1)
```

```
[90] # Label encoding
     from sklearn.preprocessing import LabelEncoder
     labelencoder = LabelEncoder()
     columns_encoding=['Station Name','Destination Station Name']
     for i,column in enumerate(columns_encoding):
         X_train[column]=labelencoder.fit_transform(X_train[column])
         X_test[column] = labelencoder.fit_transform(X_test[column])
```

- Here, we divided our dataset prior to label encoding to avoid data leakage, ensuring the integrity of our model's performance.
- Only the 'Station Name' and 'Destination Station Name' columns were subjected to label encoding, as they were deemed relevant. Other columns were dropped as they were considered less pertinent to the analysis.

Model training- [Linear regression in-depth]

Model training

```
[96] from sklearn.linear_model import LogisticRegression
     lr = LogisticRegression()
     lr.fit(X_train,y_train)
```

```
/usr/local/lib/python3.10/dist-packages/sklearn/utils/validation.py:1143: DataConversionWarning: A column-vector y was passed when a 1d array was expecte
  y = column_or_1d(y, warn=True)
▾ LogisticRegression
LogisticRegression()
```

- Ultimately, we are employing the Logistic Regression linear algorithm to predict whether a passenger is willing to share their berth or not.

Accuracy test–

```
[100] # accuracy on training dataset
      from sklearn.metrics import accuracy_score
      accuracy_score(y_train,y_pred_train)

      1.0
```

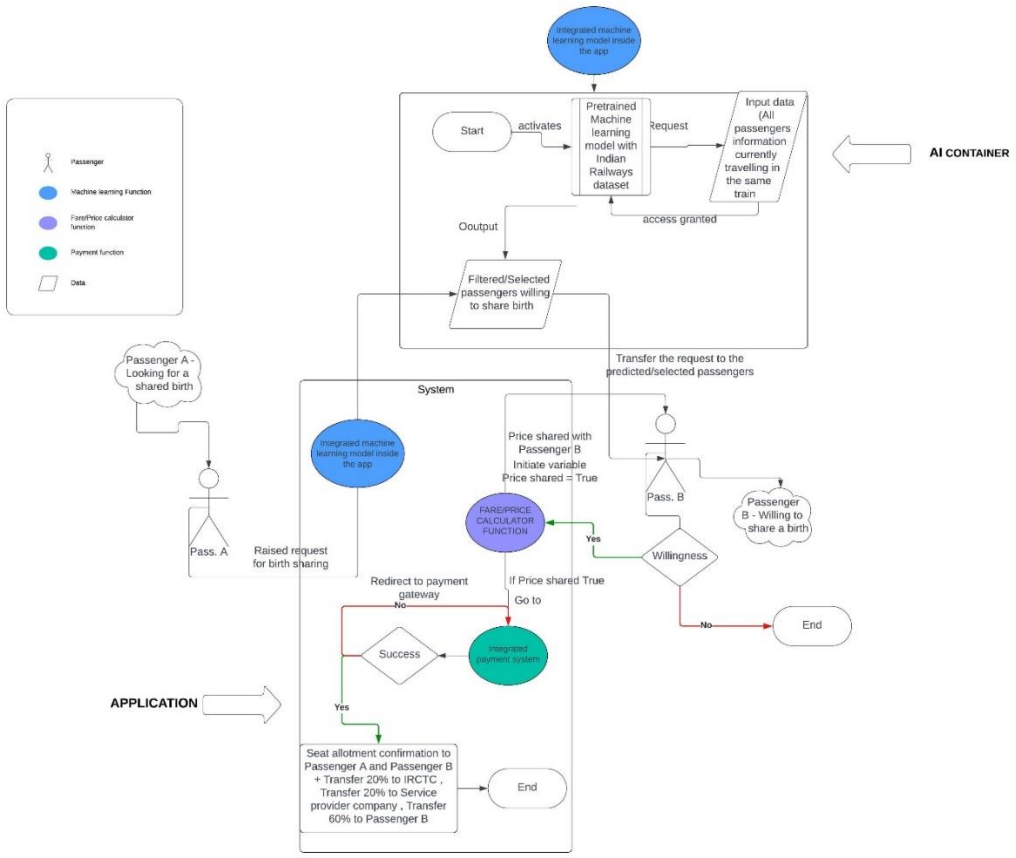
```
[102] # accuracy on test dataset
      y_pred_test = lr.predict(X_test)
      accuracy_score(y_test,y_pred_test)

      1.0
```

- As expected, the reason for having 100% accuracy in train and test dataset is having only 1 independent feature contribution.
- Since we just dumped target feature's values based on only one feature (night_travelling) because of that our model did not find any trouble in finding the pattern
- So whatever value is given in input (night_travelling), same value is returned in output (Willing_to_share).

IV) Business modelling –

Flow chart-



Model working:

Our model is trained on a comprehensive dataset provided by Indian Railways, the details of which can be found in the model prototype documentation. Let's consider a scenario where Passenger A requests to share their berth. The request is then forwarded to the model for processing.

The model utilizes the passenger information, represented by the X_test dataset, to predict the willingness of other passengers to share their berth. Subsequently, notifications are sent to all passengers identified by the model as potentially willing to share their berth.

Upon receiving the notification, the receiver passenger has the opportunity to review Passenger A's profile. If the receiver passenger agrees to share their berth, they accept the request. At this point, a fare/cost calculation function is invoked.

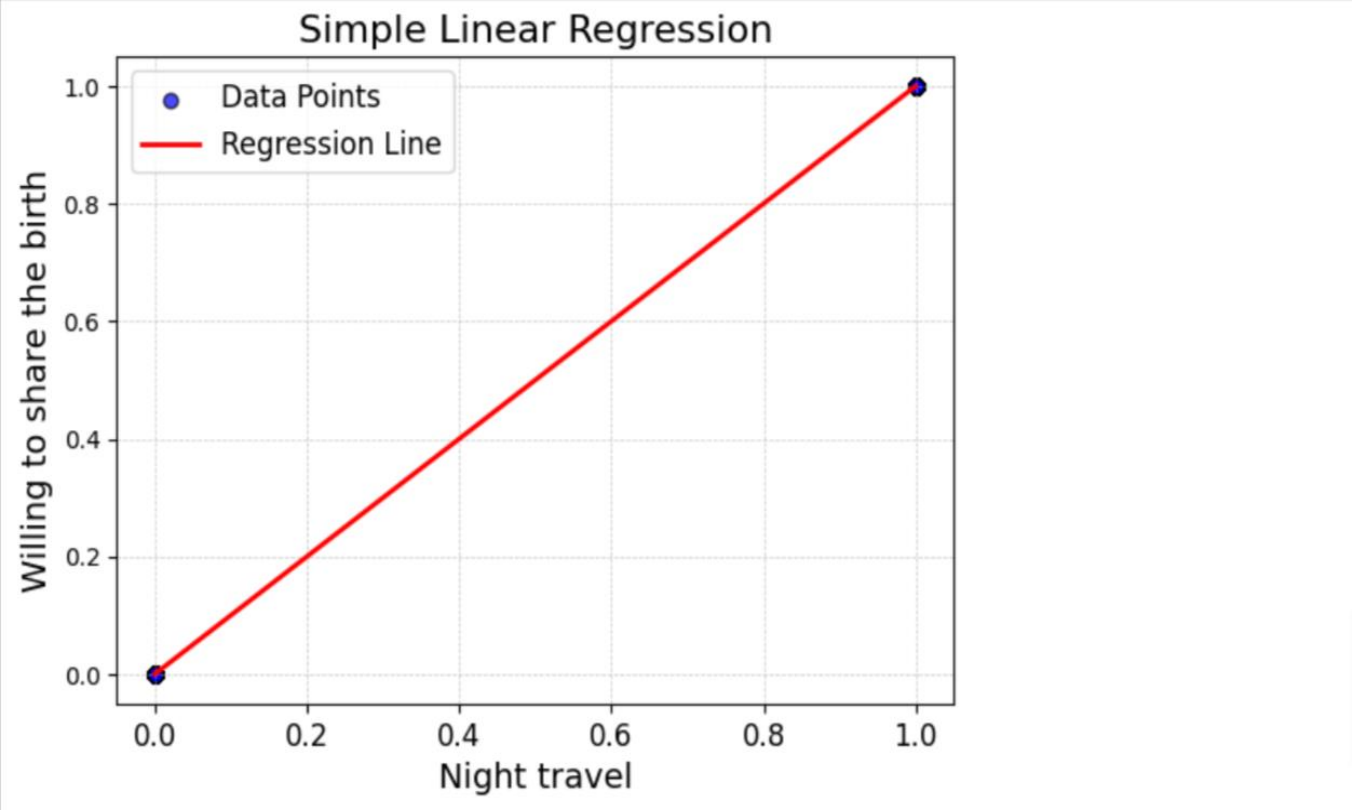
Once the total amount is calculated, it is communicated to Passenger A, who initiated the request. If Passenger A accepts the fare, an integrated payment function is triggered, which interfaces with the payment gateway to facilitate the payment transaction.

Following the successful payment transaction by Passenger A, the total amount is split among Indian Railways (20%), the service provider (20%), and Passenger B (60%).

Finally, Passenger A receives a notification containing the coach number and seat number, enabling them to share the berth with Passenger B.

V) Financial modeling (equation) with machine learning and data analytics-

[Follow this link to understand more about regression line**]**



- As depicted, both lines are entirely overlapping, indicating that our model merely replicates the input feature, which is 'Night travelling'.

Business idea market:

Market Analysis: The market potential for AI-driven berth sharing in Indian Railways is substantial, considering the sheer volume of passengers utilizing the rail network daily. With over 23 million passengers traveling on Indian Railways every day, there exists a vast untapped market for services that enhance passenger comfort and affordability. Additionally, factors such as increasing urbanization, rising disposable incomes, and growing digital adoption further contribute to the attractiveness of this market.

VI) GitHub link-

https://github.com/codedestructed007/FeyNN-labs-Internship/blob/main/Indian_railway_prototype_dev.ipynb

VII) References –

1. https://timesofindia.indiatimes.com/education/learning-with-toi/indian-railways-10-surprising-facts-that-every-student-should-know-about/articleshow/105430676.cms
2. https://www.graphpad.com/quickcalcs/linear1/#:~:text=The%20formula%20for%20simple%20linear,b%20is%20the%20estimated%20intercept.
3. https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/
4. https://www.analyticsvidhya.com/blog/2021/10/everything-you-need-to-know-about-linear-regression/
5. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html
6. https://economictimes.indiatimes.com/industry/transportation/railways/three-years-on-indian-railways-passenger-traffic-remains-below-pre-pandemic-numbers/articleshow/99458137.cms?from=mdr