Capstone Project Proposal

Amisha Muni

Business Goals

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you're labeling images, how will this help the business?	*General background: This is written from the perspective of the industry I currently work in – which is a company developing Software for Airlines. Airlines often request us to propose solutions to specific situations they are trying to address.* A large airline operates many flights out of London Heathrow and has invested a lot of money in offering excellent Lounge facilities, to passengers who fly in Business or First class. The Lounges are open from 05:00 till 22:00 daily, and from the usage data, the airline can see that there are periods of time during which the Lounges are under-utilized. The airline would like to find a way to monetize this unused Lounge capacity by selling discounted Lounge access to passengers travelling in Economy class, but in a way that is well thought out, and controlled so that it doesn't result in overcrowding in the Lounge. The system has to adapt to changing passenger traffic. The goal of the project is to investigate an ML/AI solution for predicting periods of Lounge under-utilization, and constantly improving those predictions based on actual passenger traffic. The predictions of the model can then be incorporated in the passenger check-in flow of the airline, so that an appropriate number of Economy passengers who are likely to be interested in purchasing discounted Lounge access are sent a real time offer soon as they check-in.

	The main parts foreseen for this project are: Predict Lounge under-utilization by building a model of the Lounge usage based on: - Actual Lounge usage production data from the Airline - Flight schedules & inventories, to know the traffic for the Airlines flights on a daily basis, and the booked Economy, Business and First class passengers. - Passenger data from the Airline - Restrict scope to the London Heathrow lounges to start with. Feedback loop to correct the predictions made Prove model accuracy for the London Heathrow lounges, so that it helps the Airline build a business case to sell discounted Lounge access, and finally rollout both developments worldwide. **Strict access control and anonymization of all data given to us by the Airline needs to be put in place, as per GDPR regulations.** The ability to fully delete any passenger specific data on request by a passenger needs to be developed, as per GDPR regulations. Full explainability of proposals made by the model is required, so that any potential bias around nationalities, genders etc. can be identified and removed. A future development foreseen is: Mechanism to offer discounted Lounge access to a suitable number of Economy class passengers for the under-utilized periods of time identified, via the passenger check-in flow. This will then result in revenue for the airline when passengers buy this access.

Business Case Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.	Airlines across the world invest a lot of money in building and offering Lounge access to their First and Business class passengers. It is almost an unwritten expectation that Lounge access is included for First and Business class tickets. However what is also true is that the number of First and Business class tickets sold is a fraction of the Economy class tickets sold. There is growing financial uncertainty in the world with various political events, and this is impacting the premium end of the market more. Lounges stand quite under-utilized at various times during their operating hours, and there is a cost to keep them open. There is also a growing middle class who wont purchase an expensive Business or First class ticket, but will be more open to purchasing an ancilliary service like Lounge access. The business value for the airline here is on the following aspects, **once the Lounge utilization can be accurately predicted**: - Sell discounted access to the Lounge to Economy class passengers, and actually monetize the under-utilized capacity of the Lounges - Expose a new segment – the Economy travellers to the Lounges, thereby hoping to convert a few of them to purchasing a Business class ticket in the future. Or just towards them enjoying their journey more and increasing customer satisfaction – people are more relaxed getting on a flight when they are better rested in a Lounge before boarding - Have concrete data to see the actual scale of under-utilization, and an ability to vary different parameters, for example the discounted Lounge access price or the opening hours for the Lounge, to see what really strikes a good balance for the airline.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	ML/AI needs to be used in generation of models which consume a lot of different data given by the airline and learn the current Lounge utilization. Then when subjected to fresh data, the models need to be able to propose the level of Lounge under-utilization, and the periods of time there is likely to be under-utilization over.

	Usage of the predictions from the model: This data can then be integrated in the check-in flow for the Airline, so that soon as a person checks in, the system uses the predictions of the model to know if it should aim to offer the passenger a discounted Lounge access, and for which Lounge. The check-in flow can also be enhanced for Business and First class passengers by suggesting which of the Lounges they could use, when an Airport has more than one Lounge.

Success Metrics

Success Metrics What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.	Proposed Metrics: 1. Predicted Lounge occupancy, per Lounge 2. Passenger figures by class & destination, at Airport 3. Actual Lounge occupancy & method of Lounge access (e.g. purchased at discounted price, or included in ticket fare) Baseline values for points 1-3 can be obtained from current data supplied by the Airline. Note today there is no discounted price Lounge access sale. Discounted price Lounge access sale will not be available till further changes are done to integrate the predictions and offers to Economy class passengers in the check-in flow.

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	The Airline has a contractual agreement with the Passenger who is flying on a flight booked with them. This fine print grants the Airline ownership of the data given by the customer – e.g. payment mechanism, booking data, tickets etc., to hold for a period of time for legal reasons. In line with GDPR, the data may only be used for the purpose for which it is collected, and only by the Airline which is authorized by the passenger (the legal contract is by buying the ticket to fly on a flight operated by the airline). The Airline has the rights to grant our company access to this data, to develop a solution for the airline. **The data we are granted access to cannot be used by us for any other purposes, and access to it has to be strictly controlled and anonymized.** We need to also develop the ability to fully delete any copies of this data if a Passenger requests deletion by the Airline, as the Airline will pass this request to us. Airlines maintain this data already in warehouses, so a large amount of data will be available to us, and we will need to isolate the aspects we need. Secure Physical connectivity with the Airline warehouse will need to be established (e.g. secure MQLink) to receive this information. **As this is a request from the Airline, we do not have to pay for usage of this data, however a detailed statement of estimated costs for the hardware and software need to be given to the Airline for consideration for the approval of this development.** The model developed will need to have a feedback loop to constantly improve its predictions based on ongoing production data, as flights are regular and scheduled and every flight has a different passenger load factor – there can also be seasonal variations, on top of natural emergencies (e.g. weather based disruption or a terrorism incident) The initial development should be a prototype with the scope being Lounges for the Airline at London

	Heathrow.
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	The following aspects of data are needed: - Airline Schedule data - Airline Booking Inventory data - Airline Lounge access data Key elements of data: - Airline Schedule: Flight number, departure date & time, destination - Airline Booking and Inventory: Number of passengers booked in each class, on each Flight - Passenger entry and exit from Lounge, date & time, passenger flight number, passenger class By the nature of this data, there are items that are strongly restricted and protected under GDPR. For the prototype, we propose that the past 2 years of Airline data for London Heathrow is used for training, validating and testing the model. The anticipated size of the data is 1GB a day. <u>Potential biases built into the data:</u> - There could be large spikes in Lounge occupancy at times, when there is a Disruption or Delay to many flights. - There could be large dips in Lounge occupancy at times, when there are mass cancellations of flights, say due to a planned strike or adverse conditions in a destination country. - Seasonal variations in passenger data are to be expected – e.g. more families travel in school holiday periods, there can be more travel to a specific region around festivals - An airline can announce a new route, leading to new passenger traffic - An airline can purchase a new aircraft which is larger, leading to more passenger capacity on some flights - Passengers may start being more conscious about the environment and fly less, leading to less bookings on some routes
Choice of Data Labels What labels did you decide to add to your data? And why did	The following labels are proposed: - Very Low: <20% - Low: 20-40%

you decide on these labels versus any other option?	- Medium: 40-60% - Medium High: 60-70% - High: 70-80% - Very High: 80-100% The purpose of the model is to predict the Lounge occupancy – i.e. Lounge utilization, and therefore given data about upcoming flights and passenger bookings, the model needs to predict how busy each Lounge is going to be over periods of time. This then will allow the airline to know how many passengers they can try to sell discounted Lounge access to.

Model

Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	To build an accurate model (or models) and identify and extract the relevant data needed to train and test it, it is vital to use inhouse resources we already have, as they have both the functional knowledge of the complex Airline industry domain, as well as the technical knowledge and a platform to build the solution on. The data we are passed has a strong GDPR component, and this further restricts our ability to use external platforms. Additionally, we can easily integrate a real-time check to the model's prediction in the passenger check-in flow for the Airline in future, as we already host it. This integration should be done after the accuracy of the model is established and accepted by the airline and is indicated in more detail in the roll out plan.
Evaluating Results Which model performance metrics are appropriate to measure the success of your	The model needs to predict the Lounge utilization for each day, for every hour that the Lounge is open for. The expectation on the prediction is quite high, as if the Lounge gets overcrowded because it gets oversold later based on the predictions, then it may result in

model? What level of performance is required?	complaints from passengers. We target a F1 score of 0.9 Confusion matrix should be used to check that we have a balanced spread of data in what is received from the airline. We need high precision and high recall, and given a choice between the two, higher precision is preferred. i.e. when the model does predict the level of Lounge utilization, it should be highly accurate and err on the side of caution. e.g. when in doubt an Airline would sell less Lounge access rather than oversell.

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	**The scope of the MVP is:** - Lounge utilization modelled for London Heathrow Lounges for the airline - Training, Validation & Testing dataset is the data for previous 2 years from the airline - Metrics logged to allow an indication of Lounge utilization - Run the trained model as a prototype in production for 3 months, to prove its accuracy. - Feedback loop for improvements to model Closure of MVP will be defined by model accuracy acceptance by airline, and progress to integration of the model in the existing passenger check-in flow.
Use Cases What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will	*I prefer to structure & track my product vision via Initiatives and Epics, where Initiatives in JIRA cover multiple related Epics.* **Persona:** The user is an Airline, which wants a view of the

users access this product?	predicted utilization of its Lounges at London Heathrow, based on its passenger bookings and flight schedules. **<u>Initiatives:</u>** 1. **Data Acquisition:** This initiative would gather all the Epics needed for actual Link setup between the Airline and the company, and also validation of reception of necessary data from the Airline. 2. **Data Ingestion and Extraction:** This initiative would gather all the Epics needed to ensure all the types of data are well ingested, relevant information is extracted, transformed into the format needed for the models, and stored. There can also be work to filter spikes in the data, i.e. the potential bias points listed in earlier sections. 3. **Lounge Utilization Modelling:** This initiative would gather all the Epics needed for creating, validating and testing various models suitable for arriving at predictions of the Lounge occupancy. We should also see how the model behaves with data related to common scenarios – e.g. a Disruption scenario, when passengers either don't arrive or stay in the lounge for longer, as their flight is cancelled or delayed. Or change in passenger traffic as a new route is added or new larger aircraft purchased. 4. **Metrics:** This initiative would gather all the Epics needed for analyzing and logging all the Business as well as Technical metrics. Business metrics include the 3 points listed in an earlier section. Technical metrics include any metrics needed to monitor the platform and processing. 5. **Go Live (3-month evaluation in Production):** This initiative would gather all the Epics to actually deploy and evaluate the predictions of the model on Production data, after the model is trained. There would also need to be a feedback loop so that the model improves based on the new data it gets exposed to. We may even have more than one model, and this can help narrow down the most suitable model. 6. **MVP Review & Closure:** This initiative would gather all the Epics to firstly present to the airline the results of the model predictions for London

Heathrow lounge utilization, versus the actual lounge utilization observed over a 3 month period.
The final step would be to get Airline approval for the model, and to then decide the next steps (e.g. around model extension to other airports and helping the airline use the proven predictions in its check-in flows, estimation of cost)

Epics:

Initiative: **Data Acquisition**
Link Setup (to receive data from the Airline)
Validate reception of Airline Schedule Data
Validate reception of Airline Booking & Inventory Data
Validate reception of Airline Lounge Access Data
Initiative: **Data Ingestion and Extraction**
Extract, Transform, Store: Airline Schedule Data
Extract, Transform, Store: Airline Booking & Inventory Data
Extract, Transform, Store: Airline Lounge Access data
Initiative: **Lounge Utilization Modelling**
Investigate and build suitable models
Validate models
Test models
Initiative: **Metrics**
Analyze and log functional metrics
Analyze and log technical metrics
Initiative: **Go Live in Production**
Deploy models in Production
Feedback loop to adjust model Predictions
Initiative: **MVP Review**
Present model prediction vs actual occupancy
Closure and next step agreement

Roll-out

How will this be adopted? What does the go-to-market plan look like?

Roll-out and go-to-market plan:

1. Build, train and deploy model using dataset from last 2 years, for London Heathrow only.
2. Run the trained model for 3 months in production with a daily comparison of its predictions against London Heathrow lounge actuals.
3. Adjust model on a regular basis if needed.
4. Present the summary of model predictions of lounge utilization vs actual lounge utilization for the London Heathrow lounges, for this 3 month

	period, to the Airline. 5. Post airline approval, decide the next steps jointly with the airline. 5.1 Integration of model predictions in the check-in flow, so that Airline can start to sell excess Lounge capacity to Economy class passengers from London Heathrow, and monitor their revenue 5.2 Extend the model training dataset to include data from other airports for the airline, so that all changes can be rolled out to more and more airports and Lounges used by the airline. 5.3 Pricing for these changes The reason we go very gradually is that unless the accuracy of the model is proven, the Airline will not sell Lounge access to Economy passengers. Once the model accuracy is acceptable, and there is revenue flowing in from the sale of Lounge access, the airline will then be in a better position to fund deployment of this solution to their whole network.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	The real world for an Airline is very complex and things can change a lot from day to day. e.g. a strike or a volcano eruption can cause a lot of flights to be cancelled from the departure airport, weather conditions at an arrival airport can cause a delay to departure, holidays and seasons result in change number of passenger bookings, an airline can add a new route which impacts the number of passenger bookings, people may prefer to fly less in certain regions leading to drop in bookings, an airline can use a larger or smaller aircraft on a route While we will train our model with previous production

	data received from the airline, it will never be a perfect representation of actual production data – that will really change from day to day, and its for us to model and train so that we take into account the main factors and aim to have a generally good prediction. A/B testing can definitely be used, to evaluate different models and see which one performs best. We may even find we need a range of models and then use the decision that most of them have a consensus on. We have to constantly use a feedback loop and learn from the actual Lounge occupancy at the end of the day. We have to analyze cases where our predictions are incorrect. Regular interaction with airline industry experts will help in ensuring we keep evolving the model in an accurate way and keeping it up to date. The airline industry changes quite rapidly. A future integration of the predictions in the check-in flow can also give rise to models which help the airline to learn better which passengers actually purchase Lounge access based on the offer sent to them, and which are less likely to purchase the same. **For these future models we will have to be extremely careful to avoid additional unwanted bias** around Gender, Destination, Nationality or Passenger booking types (e.g. Single travelers vs Families with children, women vs men, certain Nationalities rated more favorable for a purchase lounge access offer, passengers to certain destinations rated less suitable for a purchase lounge access offer)
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	In Lounge utilization predictions, we can get unwanted bias due to positive and negative spikes which happen due to real incidents. e.g. Strikes / cancellations / delays can lead to less or more lounge utilization. Passenger bookings can go up or down in different classes Airlines can add more routes or use smaller or larger aircraft. We should be careful that we consider these well in our simulations and avoid biasing predictions on these situations.

	Continuous evaluation of the model's predictions against actual Lounge utilization will be a good way to ensure accuracy and avoid bias. Updating and retraining the model on a regular basis will also ensure accuracy and avoid bias, as the real-world situations will keep changing. Input from business experts in the airline will be key to this regular updating and retraining.