

Project Progress Report 2

(due May 24th 11:59p.m)

You can start working on the project once your report is accepted and graded by your TA. The entire final project is worth **35%** of your final grade and this report accounts for **10%**. This project is done individually.

Submission Guideline

Download this google doc, fill the table. **Type** your answers, no handwritten answers will be accepted (except for the very last question). Submit it in **PDF** format on Gradescope.

If you need some inspirations please feel free to take a look at:

[Showcase of Information is Beautiful Awards](#)

[Bloomberg Year In Graphics Review](#)

[The Pudding](#)

[The New York Times](#)

Project Guidelines

Note: The guideline has been further clarified from Progress Report 1, so double-check whether your dataset choice still satisfies the updated guideline below.

1. You may use more than one dataset, however, regardless if you use one or multiple datasets, your visualizations must make use of at least three following data types - **link, position, and attribute**.
2. You cannot use any dataset from the class (Labs, Assignments, Lecture Exercises)
3. You can make your own dataset (Web scrape etc.) provided point 1. is satisfied.

Part 1 - Story and Narrative

Link to the dataset	https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction
Example item from the dataset	0 11-02-2022 SpiceJet SG 8709 18:55 Delhi 02h 10m non-stop 21:05 Mumbai 5,953
Story you want to deliver	<p>(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide)</p> <p>Facts:</p> <ul style="list-style-type: none">• Business class seats are likely to cost more than economy class seats• Air India, Vistara, SpiceJet, AirAsia, GO FIRST, Indigo, Trujet, StarAir are the 8 most popular airlines in India• Air India and Vistara are the only airlines that offer business class flights• Data provided contains flight data between 6 cities: Bangalore, Chennai, Delhi, Kolkata, Mumbai, and Hyderabad• From these cities mentioned above, the most popular cities as of 2022 were: Mumbai, Delhi, Bangalore, Hyderabad, Chennai, and then Kolkata• Bangalore and Chennai are really close together on the map <p>Insights:</p> <ul style="list-style-type: none">• Most profitable airlines are Vistara and Air India• People who choose to fly long distances usually start flying in the morning• Passengers choose to arrive at their destinations at night• Indian airports are busiest during two times of the day: morning and night• Air traffic is closely related to the population size of each city• Delhi is a more popular destination than Mumbai even though it has a smaller population• Delhi and Mumbai are the most traveled to places <p>Messages:</p> <ul style="list-style-type: none">• Airlines with name recognition are more likely to be the most profitable. This may take time. Deploy advertisements while targeting populous cities and avoid scandalous coverage

	<ul style="list-style-type: none"> ● Morning flights should be more expensive as people who take long flights prefer to do so in the morning. Pricing should strike a balance ensuring that customers do not switch to fly with another airline ● Flights between big cities are most popular ● Make flights available during busy hours ● Maintain active flight operations in populous cities such as Mumbai, Delhi, and Bangalore ● Investigate the reasons why customers choose Delhi and Mumbai over other cities. Find a way to promote these less popular cities
Describe your target audience.	<p>My target audience are Indian airline executives who are curious about the trends and demand for flights in India's most populous metropolitan cities. I can also use this data as a case study if I am assisting airline executives in another country such as the United States since the trends in data should be the same.</p> <p>Familiarity with your topic? If not, how do we catch them up? Yes, Indian airline executives are domain experts which means they will be familiar with my topic.</p> <p>Do they care? Why? Why not? Yes, my audience do care as they want to stay competitive in the airline industry.</p> <p>What do you want them to take away? Key points?</p> <ul style="list-style-type: none"> ● The most popular routes and destinations ● The most popular time of days for traveling ● Most profitable airlines, and the characteristics of these airlines ● The most popular destinations <p>What do they know about visualization? Are your techniques standard? As domain experts, my target audience is likely to understand simple data visualization techniques such as line, bar, and pie charts. If I plan on making an interactive or dynamic</p>

	<p>visualization, however, I will need to ensure my visualization is simple and intuitive to ensure that it is properly interpreted.</p> <p>How do they encounter your visualization? I would imagine this visualization to be encountered during a meeting/conference to discuss airfare trends and insights.</p> <p>Mathematical background? Are you assuming too much? Too little? Because the role of my target audience involves making decisions based on data, I would assume Indian airline executives have a decent mathematical background. If necessary, a brief explanation of any statistical or mathematical concept may be provided to ensure the visualization is properly interpreted.</p> <p>Device? Mobile phone, computer, print media... My target audience will likely encounter my visualization on computers or laptops. These devices have larger screens which will allow for small details in my visualization to be visible.</p>
The goal of your project outcome. And why?	<p>(exploratory vs. explanatory)</p> <p>My goal is to create a visualization that will provide valuable insights on the trends found in itinerary data pertaining to India's most populous metropolitan cities. Because my audience are domain experts who care most about valuable information, my visualization will be more exploratory than explanatory as I will be putting aesthetics second to the information I convey.</p>
Narrative structure you plan to use	<p>Linear interactive slideshow</p>
Elaborate your choice of narrative structure.	<p>I've chosen linear structure since my data is pretty straightforward and prefer telling how my data is rather than having my audience explore it. This will hopefully deter any miscommunication.</p> <p>each slide will be for each visualization: when, where, and why how passengers fly. It is interactive since users can zoom into the visualizations and hover to get a closer look. All of my</p>

	visualizations also have tooltips so that the user can get take a look at the numerical data
Narrative genre you plan to use	Annotated chart
Elaborate your choice of narrative genre.	I want to do an annotated chart as it is clear, intuitive, and allows my audience to save a lot of time.

Part 2 - Outline

Story you want to deliver	<p>(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide)</p> <p>Facts:</p> <ul style="list-style-type: none"> ● Business class seats are likely to cost more than economy class seats ● Air India, Vistara, SpiceJet, AirAsia, GO FIRST, Indigo, Trujet, StarAir are the 8 most popular airlines in India ● Air India and Vistara are the only airlines that offer business class flights ● Data provided contains flight data between 6 cities: Bangalore, Chennai, Delhi, Kolkata, Mumbai, and Hyderabad ● From these cities mentioned above, the most popular cities as of 2022 were: Mumbai, Delhi, Bangalore, Hyderabad, Chennai, and then Kolkata ● Bangalore and Chennai are really close together on the map <p>Insights:</p> <ul style="list-style-type: none"> ● Most profitable airlines are Vistara and Air India ● People who choose to fly long distances usually start flying in the morning ● Passengers choose to arrive at their destinations at night ● Indian airports are busiest during two times of the day: morning and night ● Air traffic is closely related to the population size of each city
---------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

	<ul style="list-style-type: none"> • Delhi is a more popular destination than Mumbai even though it has a smaller population • Delhi and Mumbai are the most traveled to places <p>Messages:</p> <ul style="list-style-type: none"> • Airlines with name recognition are more likely to be the most profitable. This may take time. Deploy advertisements while targeting populous cities and avoid scandalous coverage • Morning flights should be more expensive as people who take long flights prefer to do so in the morning. Pricing should strike a balance ensuring that customers do not switch to fly with another airline • Flights between big cities are most popular • Make flights available during busy hours • Maintain active flight operations in populous cities such as Mumbai, Delhi, and Bangalore • Investigate the reasons why customers choose Delhi and Mumbai over other cities. Find a way to promote these less popular cities
<p>Specifications on each plot in the order of how you lay out on your project</p>	<p>(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story)</p> <p>1. Plot 1</p> <ol style="list-style-type: none"> 1) Task: This chart a) analyzes trend between flight time and flight duration and number of flights b) determines whether flight duration or flight time plays a role in class choice determines the busiest times for departure and arrival 2) Attributes: departure time, period (self-derived attribute), duration, class (self-derived attribute), arrival time 3) Marks: point mark, line mark, bar mark 4) Channels: <ul style="list-style-type: none"> - aligned vertical position channel for duration number of flights (aggregate) - aligned horizontal channel for departure time - color channel for class departure time - color channel for period arrival time

—shapes channel for period

5) How this plot adds to the story:

~~My visualization aims to uncover two of the factors that lead to class choice. Identifying when passengers fly is also important because it allows airlines to adjust pricing to maximize profitability. Essentially, this plot will answer the “when” part of passenger itinerary decision-making.~~

Uncovers the busiest times of the day (The “when” aspect of flight management)

2. Plot 2

1) Task: This chart determines popular routes and destinations

2) Attributes: from (origin city), to (destination city), airline

3) Marks: nodes mark, line mark, text mark

4) Channels:

- connection channel between each origin and destination city
- line thickness to indicate route popularity
- size channel to differentiate city population

5) How this plot adds to the story:

My visualization aims to identify the most popular routes and destinations for pricing purposes. This plot will allow my audience to uncover connections between city size and route/destination popularity. Essentially, this plot will answer the “~~where~~ how” part of passenger itinerary decision-making.

Plot 3

6) Task: ~~This chart compares the average number of flights with a flight on Indian Independence day~~

This chart uncovers the most popular destinations

7) Attributes: number of flights (aggregate), destination city

	<p>8) Marks: bar mark, area mark, text mark</p> <p>9) Channels:</p> <ul style="list-style-type: none"> - Size channel for number of flights — Color channel comparing the number of flights on average day vs flight on Indian Independence Day - Text channel outlining number of flights - Position channel to show where the city is on the map - Color channel of each circle <p>10) How this plot adds to the story:</p> <p>My visualization aims to compare the average number of flights of each airline compared to the number of flights on Indian Independence Day. This is to give an idea of how much more busier the airline will be on this day.</p> <p>My visualization aims to make any inferences between the popularity of destination cities with where the city is located. This is the “where” aspect</p>
<p>Elaborate the choice of their marks and channels for each vis</p>	<p>1st vis:</p> <p>point mark: each data point represents a specific flight (combination of flight period, duration, and class). I chose point marks to see if I can identify any clusters within my data. These clusters should be where airlines should target customers.</p> <p>Line mark: I made it white against the blue background so that it stands out.</p> <p>Bar mark: I was going to make it side by side, but I found overlapped gave a better comparison between departure and arrival</p> <p>aligned vertical position channel: I chose flight duration because it is the predicted target; flight time period predicts duration.</p>

~~aligned horizontal position channel: I chose period as the x-axis because it is the predictor; flight time period predicts duration.~~

aligned vertical position channel (Flights): I chose total number of flights instead of average since outliers may skew the data

Aligned horizontal position channel (time): I originally chose period, but I felt like that didn't give the full picture Now I use time.

~~color channel (class): I chose to use color to indicate class as duration and period can be used to predict class. (Note that since Air India and Vistara are the only two airlines which have business class, they will be the only airlines included in this scatterplot). I've chosen to use blue for economy class and red for business class. I've chosen these two colors because they are contrasting.~~

~~color channel (period): I chose to use color (grey, yellow) to indicate period to accompany departure time since period is derived from departure time. I've chose to use grey instead of black since it will make the points more visible~~

~~shape channel (period): I chose to use shapes (moon, sun) along with the color channel to indicate period so that the brain can better make associations~~

Color channel (departure time): I chose orange for aesthetic reasons (the color is part of the Indian flag)

Color channel (arrival time): I chose green for aesthetic reasons (the color is part of the Indian flag)

Color channel (total): I chose white for aesthetic reasons (the color is part of the Indian flag)

2nd vis:

nodes mark: each node represents a city which can either be an origin (from attribute) or destination city (to attribute). this mark will give my audience an idea of popular origin and destination cities

Line mark: Each line represents a route between an origin and destination city. this mark will give my audience know all of the routes of all these airlines, although, as domain experts, they should already be familiar with them

Text mark: Text will be used to show population size for each city and flight volume. This is to provide extra context of the cities my audience should be targeting.

connection channel: each connection represents a route between an origin and destination city. this mark will give my audience know all of the routes of all these airlines, although, as domain experts, they should already be familiar with them

line thickness channel: the line thickness of each line represents the popularity or volume of flights on each route. thicker lines indicate more popular routes. I chose to use line thickness since it is easy to for the brain to associate flight volume with line thickness

size channel: size of nodes can provide additional context and help identify possible correlation between city size and the route/destination volume

3rd vis:

~~Bar mark: I chose bar mark as opposed to pie since it makes it easier for the brain to compare sizes. Each 2 bars is grouped to represent each airline~~

~~Size channel: I'm using a size channel to identify flight volume. This makes a lot of sense since the brain associates size with volume~~

~~Color channel: I am using color channel to differentiate between the average day of an airline and the holiday. I think I will use orange for normal day and green for holiday. Orange, green, and the white background are the flag colors of India~~

Area mark (number of flights): I chose area mark for number of flights since it wa easy to compare the area between two circles

	<p>Text mark (the city): I chose text mark to indicate each city to make it clear to the viewer</p> <p>Position channel (city): I chose a position channel to give additional context to why a destination might be popular. I hypothesized that Bangalore and Chennai was not as popular as Mumbai because they were close together</p> <p>Color channel (for each circle): I chose color channel for each circle to match destination attribute color i had in visualization 1. This makes it clear for the viewer. Also is an aesthetic choice since green is a color of the Indian flag</p>
--	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Following sample answer about a single plot shows how detailed your answers to part 2 should be.

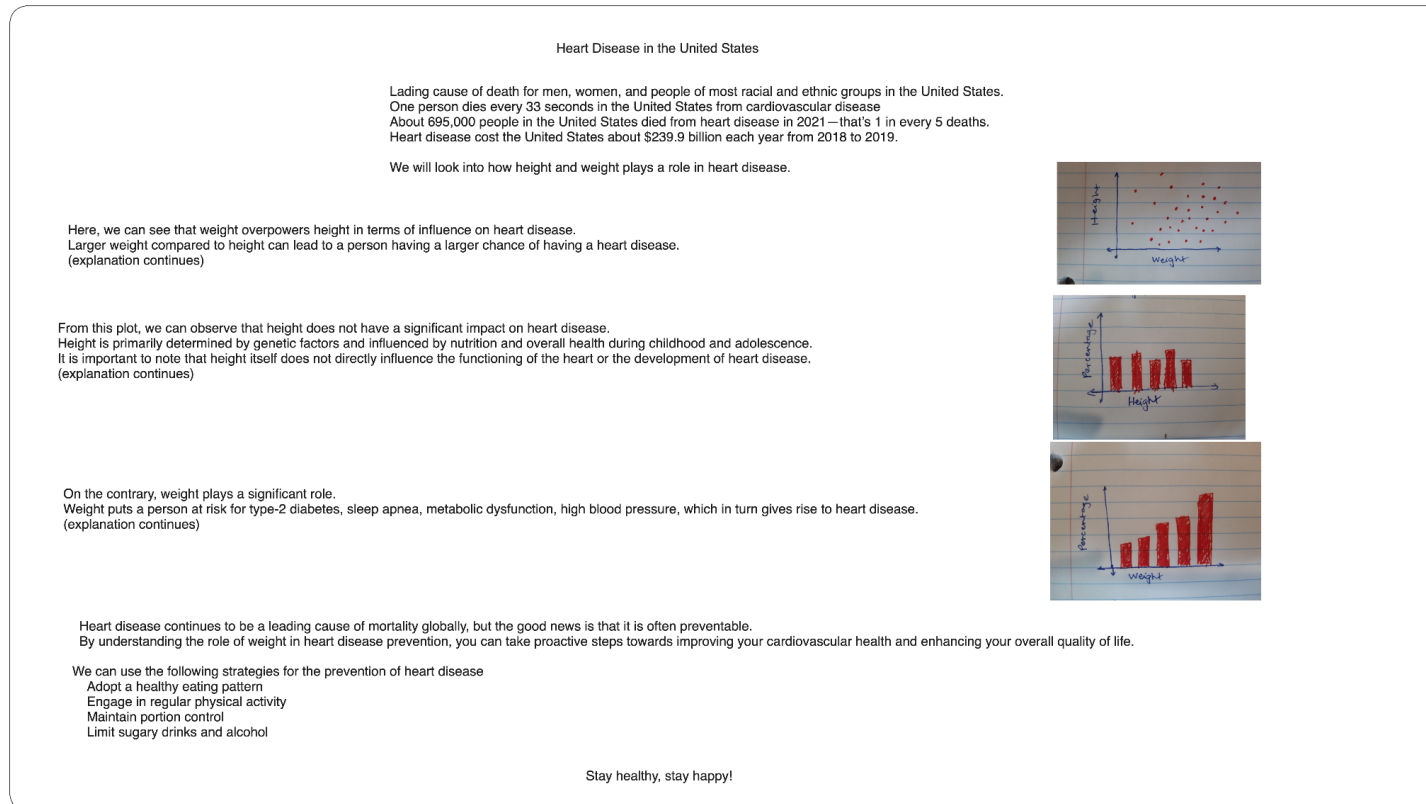
<p>1. Plot 1</p> <p>6) Task: This chart a) analyzes trend between Height and Weight of patients with heart diseases and b) locates outliers within the patients</p> <p>7) Attributes: Height, Weight</p> <p>8) Marks: point mark</p> <p>9) Channels:</p> <ul style="list-style-type: none"> - aligned vertical position channel for Height - aligned horizontal channel for Weight <p>10) How this plot adds to the story:</p> <p>My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.</p>

Part 3 - Prototype

(Look at screenshot.pdf for final product)

~~Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.~~

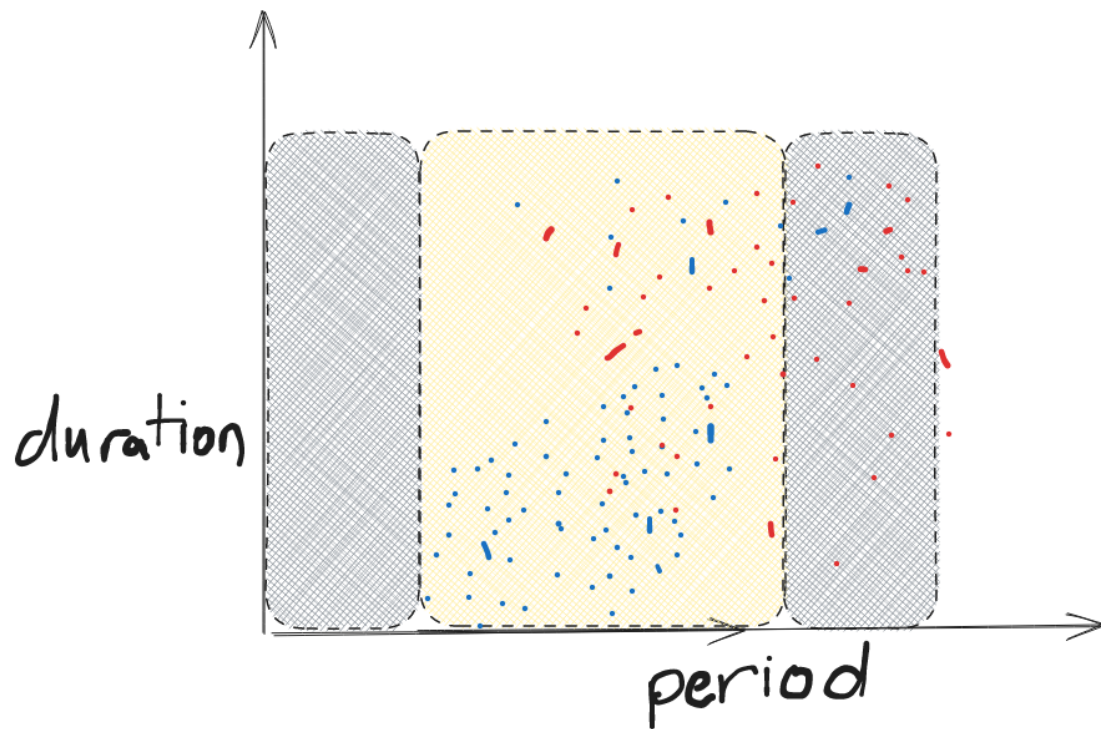
~~A basic, barebones sample prototype for this project~~



~~Flight Trends in India's Most Populous Cities~~

~~Every year, millions of passengers take to the skies to travel between Indian cities. India not only is a hub for economic activity but also a country bustling with tourists in search of the next best travel destination. With valuable insights, we can analyze trends to maximize our airline's profitability.~~

~~When do passengers usually fly? We will first analyze the relationship between flight time of day and flight duration~~



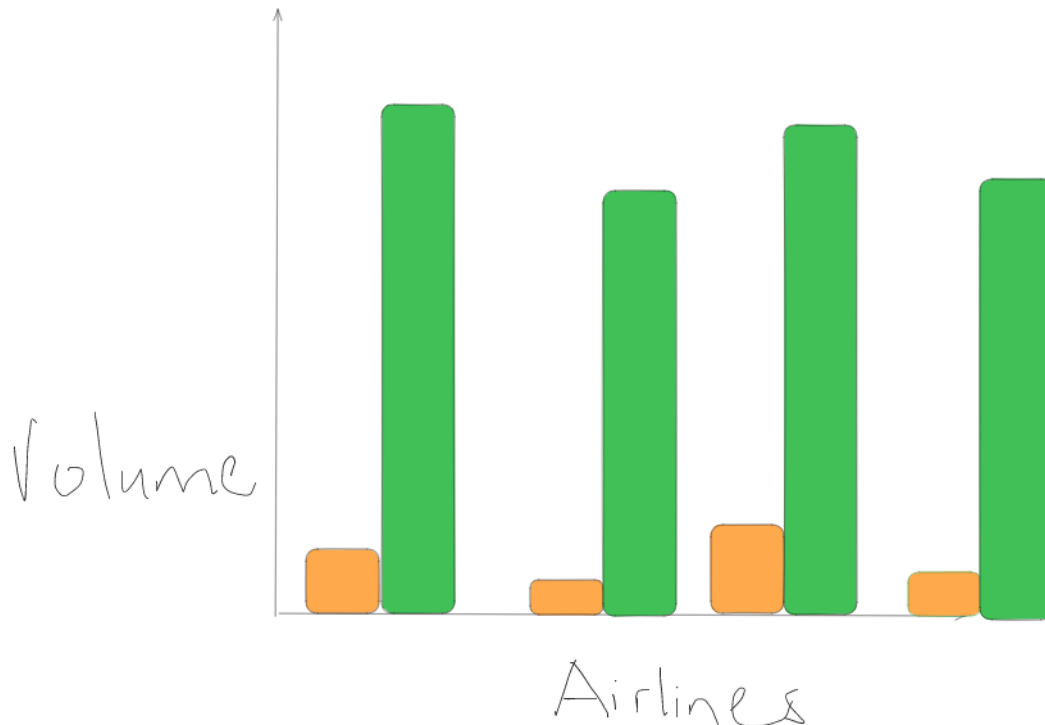
As can be seen from the visualization, it can be seen that morning flights are associated with longer flights. We speculate that this is because people tend to get up early so that they can spend the entire day traveling. Another trend that can be noted is passengers who go on longer flights tend to take business class as opposed to economy class. We suspect that this is due to passengers wanting a more comfortable flight

Now that we've analyzed WHEN passengers fly, let's take a look at WHERE passengers fly



As can be seen from the following visualization, the route between Mumbai and Delhi are the most popular. This is associated with city size since Mumbai and Delhi are the two largest cities in India.

Now we should also examine why passengers fly. Let's examine how much more busier flights are during the holiday season by examining the average number of flights vs a flight on Indian independence day.



From this visualization, we can see that on Indian independence day, airlines are busier by a large margin on Indian Independence Day. It also seems this day is busier across each airline by a proportional amount. Looks like airlines are more popular on Indian Independence Day by 57%. Airlines should increase their capacity appropriately to accommodate these passengers

From these three visualizations, we can learn several things. Since passengers are most likely to embark on long flights in the morning, it is important that we have enough of these flights to accommodate these passengers. Regarding business-class vs. economy class, we should increase the price of business class and economy class during these morning hours to maximize profits. Additionally, we should do the same with flights between populous cities because these flights will bring in the most money. Indian Independence Day has also been seen to bring in many passengers. We should capitalize on this phenomenon by increasing flight capacity as well as adjusting our rates accordingly so that we maximize our profits.