Visualisation 2

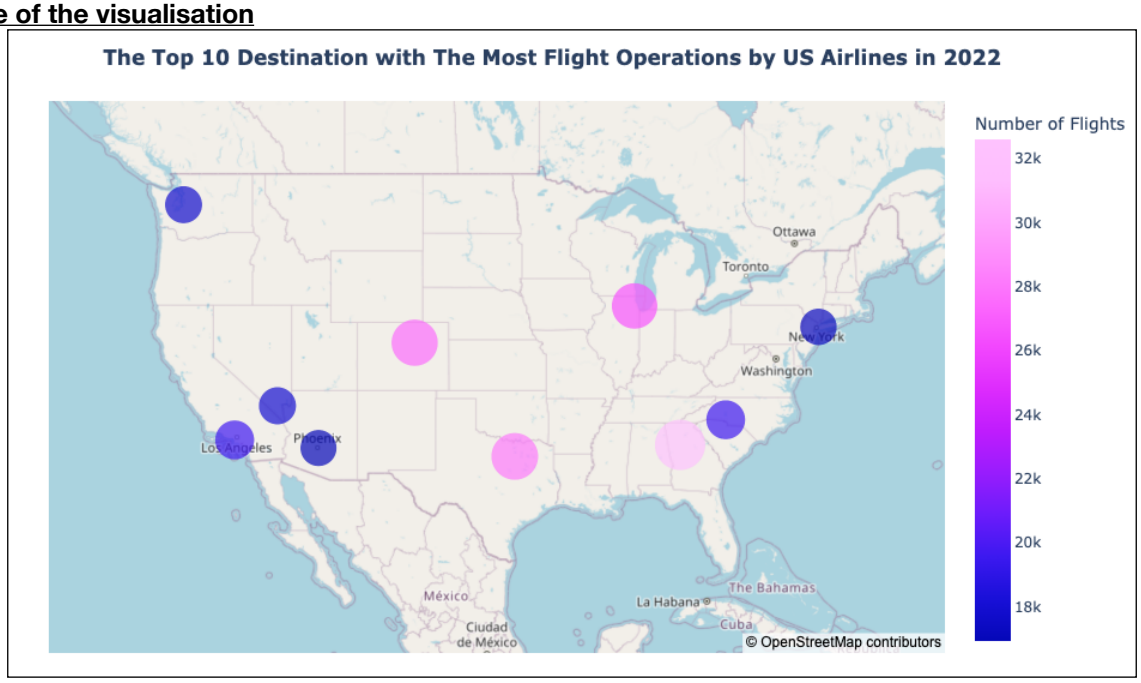
Goal

This visualisation - a bubble map chart depicts the top 10 destinations with the most flight operations by US airlines in 2022. Our target users are the business development departments of tourism-related businesses as well as the US government department of transport which works with local airports.

Many tourism-related businesses require data to investigate the next best places for their business expansion, especially near or in the airports, for example hotels, car rental companies, retailers, food & beverage and so on. Knowing the cities with the highest air traffic activities as a destination, no matter people go for travelling or business purposes, is crucial information for these companies to decide where they should pay effort to do further marketing research, evaluation and analysis.

Oh the other hand, the US department of transport, which centralises the communication with local airports, needs to identify the busiest airports with the highest number of arrival flights in order to work with these airports to review resource planning such as customs and airports staff, repair and construct airport facilities like runways and connecting bridges. These actions can ensure smooth arrivals procedures.

Image of the visualisation



Insight

As shown in the visualisation, the size and colour of the bubbles, as well as the hover labels with the exact number of flights and city names, indicate that Atlanta was a destination with the most flight operations by US airlines in 2022. It was followed by Dallas/Fort Worth, Denver, and Chicago with the pink colour range and larger bubbles. This implies that these cities may provide higher business opportunities for tourism businesses. Those with darker colours and smaller bubbles have fewer US airline flights arriving but they may still have potential (by order: Los Angeles, Charlotte, Las Vegas, Seattle, New York, Phoenix).

From the colour spectrum bar on the chart, we can discover that these 10 airports had a range of arrival flights by US airlines of around 15k-32k. With this information, the US government can allocate relevant resources or hardware support according to the traffic of a specific airport.

Data Abstraction

The raw dataset is the same as in visualisation 1, and it is preprocessed in Python to keep only the relevant columns and row (e.g. selecting 2022 because it has the latest full-year records) and to restructure the data into a new table. Also we merge another dataset to get latitude and longitude for each destination. Detailed steps will be provided in coding comments.
The abstract dataset types of the preprocessed data is geometry (2D spatial) with position data.

The abstract data type and metadata are as follows:

		Metadata (Data Semantics)	
Attribute Name	**Abstract Data Type**	**Data Type**	**Description**
Destination	Categorical key attribute	object	The city short form, a.k.a airport IATA code
City Name	Categorical (for hover label display)	object	The long form of city name of destination
Number of Flights	Quantitative sequential value attribute	int64	The number of flight operations of US airlines per year
Latitude Longitude	Quantitative diverging (for spatial position)	float64	The position of destination for marking bubbles on map

The visual encodings are as follows:

Mark
Point

Channel	**Effectiveness & Expressiveness**
1. Position - Vertical Position - Horizontal Position	- The position of each point is defined by the latitude and longitude of cities, improving accuracy and separability
2. Area (bubble size) - Varies with the number of flights	- The area of bubble, depending on numbers of flights, enhance discriminability
3. Colour Hue (bubble colour) - Varies with the number of flights	- The colour spectrum represents the numbers of flights ranging from around 15000-32000, improving discriminability - The colour hues provide better separability for numerical differences - Colour and area of bubbles provide stronger message by offering multiple channels to represent flight numbers
3. Additional Annotation - Hover labels on each city's bubble	- The hover labels display the city name, exact number of flights, and latitude & longitude, improving discriminability

Task Abstraction

Action	**Target**	**Description**	**Respective Goal**
High-Level: Discover	Trend	To get an idea of the top 10 destinations with high air traffic in 2022	As a business development user, these help to investigate the highest potential places for business expansion
Low-Level: Compare	Distribution	To compare the number of flights among these destinations to understand their potential for development	
Low-Level: Identify	Individual attribute	To study each city's air traffic condition	Government aviation staff can use this information to allocate resources efficiently with airports

Data sources

1. https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023/data?select=flights_sample_3m.csv (The data set is csv table with 32 columns and around 3,000,000 rows)
2. https://openflights.org/data.php (Download airports.dat: https://raw.githubusercontent.com/jpatokal/openflights/master/data/airports.dat)
(Copy the dataset and save it in csv format, a table with 14 columns and around 14,000 rows)