A foremost aviation industry player with a significant presence in New York City has launched an in-depth data analysis project focused on identifying trends in flight durations in air travel. This initiative aims to delve into a wealth of data related to flight schedules and operational patterns, with the objective of optimizing flight times and enhancing the overall travel experience for passengers. As the head data analyst, you have access to rich datasets, sourced from the 'nycflights2022' collection produced by the ModernDive team. These datasets include records of flights departing from major New York City airports, including JFK (John F. Kennedy International Airport), LGA (LaGuardia Airport), and EWR (Newark Liberty International Airport), during the second half of 2022. They offer a comprehensive view of flight operations, covering various aspects such as departure and arrival times, flight paths, and airline specifics:

- `flights2022-h2.csv` contains information about each flight including

Variable	Description
`carrier`	Airline carrier code
`origin`	Origin airport (IATA code)
`dest`	Destination airport (IATA code)
`air_time`	Duration of the flight in air, in minutes

- `airlines.csv` contains information about each airline:

Variable	Description
`carrier`	Airline carrier code
`name`	Full name of the airline

- `airports.csv` provides details of airports:

Variable	Description
`faa`	FAA code of the airport
`name`	Full name of the airport

```
# Import required packages
library(dplyr)
library(readr)

# Load the data
flights <- read_csv("flights2022-h2.csv")
airlines <- read_csv("airlines.csv")
airports <- read_csv("airports.csv")
```

Hidden output

```
head(flights)
```

					sched_d...	d		sched_a...	a			
1	2022	7	1	9	2129	160	118	2312	126	B6	325	N229
2	2022	7	1	12	1940	272	315	2253	262	B6	20	N591
3	2022	7	1	21	2120	181	140	2240	180	WN	548	N865
4	2022	7	1	21	2159	142	225	21	124	B6	286	N537
5	2022	7	1	22	2140	162	310	53	137	B6	500	N923
6	2022	7	1	23	2110	193	203	2259	184	YX	955	N130

Rows: 6

```
head(airlines)
```

index	carrier	name
1	9E	Endeavor Air Inc.
2	AA	American Airlines Inc.
3	AS	Alaska Airlines Inc.
4	B6	JetBlue Airways
5	DL	Delta Air Lines Inc.
6	F9	Frontier Airlines Inc.

Rows: 6

```
head(airports)
```

		name						tzone
1	AAF	Apalachicola Regional Airport	29.7275	-85.0275	20	-5	A	America/New_York
2	AAP	Andrau Airpark	29.7225	-95.5883	79	-6	A	America/Chicago
3	ABE	Lehigh Valley International Airport	40.6521	-75.4408	393	-5	A	America/New_York
4	ABI	Abilene Regional Airport	32.4113	-99.6819	1791	-6	A	America/Chicago
5	ABL	Ambler Airport	67.1063	-157.857	334	-9	A	America/Anchorage
6	ABQ	Albuquerque International Sunport	35.0402	-106.609	5355	-7	A	America/Denver

Rows: 6

```
# flights
# airlines
# airports

data <- flights %>%
    left_join(airlines, by = "carrier") %>%
    rename(airline_name = name) %>%
    left_join(airports, by = c("dest" = "faa")) %>%
    rename(airport_name = name) %>%
    mutate(air_time_hour = (air_time/60)) %>%
    group_by(airline_name, airport_name) %>%
    summarise(avg = mean(air_time_hour), n = n(), na.rm = TRUE) %>%
    ungroup()
data
```

Hidden output

```
frequent <- data %>%
    arrange(desc(n)) %>%
    head(1)
frequent
```

	airline_name	airport_name			
1	Delta Air Lines Inc.	Hartsfield Jackson Atlanta International Airp...	Invalid...	5264	True

Rows: 1

```
longest <- data %>%
    arrange(desc(avg)) %>%
    head(1)
longest
```

	airline_name	airport_name			
1	Delta Air Lines Inc.	Daniel K Inouye International Airport	10.7167	15	True

Rows: 1

```
least <- flights %>%
    left_join(airlines, by = "carrier") %>%
    rename(airline_name = name) %>%
    left_join(airports, by = c("dest" = "faa")) %>%
    rename(airport_name = name) %>%
    mutate(air_time_hour = (air_time/60)) %>%
    filter(origin == "JFK") %>%
    group_by(airport_name) %>%
    summarise(n = n()) %>%
    arrange(n) %>%
    head(1)
least <- as.character(least[1,1])
least
```

```
'Eagle County Regional Airport'
```