

Assignment result

1) What was the highest number of people travelled in which year?

Code:

```
dataset.groupBy(col(colName: "year")) RelationalGroupedDataset
  .agg(sum(col(colName: "total_no_of_booked_seats")).as(alias: "no_of_people_travelled")) Dataset<Row>
  .orderBy(desc(columnName: "no_of_people_travelled")).limit(n: 1).show();
```

Output:

```
+---+-----+
|year|no_of_people_travelled|
+---+-----+
|2007|          176299|
+---+-----+
```

2) Identifying the highest revenue generation for which year?

Code:

```
dataset.groupBy(col(colName: "year"))
  .agg(sum(col(colName: "average_revenue_per_seat").multiply(col(colName: "total_no_of_booked_seats")))
      .cast(to: "Decimal(10,2)").as(alias: "average_revenue_per_year"))
  .orderBy(desc(columnName: "average_revenue_per_year")).limit(n: 1).show();
```

Output:

```
+---+-----+
|year|average_revenue_per_year|
+---+-----+
|2013|        66363208.71|
+---+-----+
```

3) Identifying the highest revenue generation for which year and quarter (Common group)?

Code:

```
dataset.groupBy(col(colName: "year"), col(colName: "quarter"))
  .agg(sum(col(colName: "average_revenue_per_seat").multiply(col(colName: "total_no_of_booked_seats")))
      .cast(to: "Decimal(10,2)").as(alias: "average_revenue_per_quarter"))
  .orderBy(desc(columnName: "average_revenue_per_quarter")).limit(n: 1).show();
```

Output:

```
+---+-----+-----+
|year|quarter|average_revenue_per_quarter|
+---+-----+-----+
|2014|    4|        18819408.48|
+---+-----+-----+
```

****Note:** Please find the GitHub repository (<https://github.com/bithalb2/socgen-batch3-assignment>) for details.