## Aggregating Airlines Collection

1. How many records does each airline class have? Use $project to show result as:

{ class: "Z", total: 999 }

**Query:**

db.airlines.aggregate([

{$group: {\_id: "$class", total: {$sum: 1}}},

{$project: {"class": "$\_id", \_id: 0, total: 1}}

])

**Result**:

{ "total" : 140343, "class" : "F" }

{ "total" : 23123, "class" : "L" }

{ "total" : 5683, "class" : "P" }

{ "total" : 17499, "class" : "G" }

2. What are the top 3 **destination cities** **outside** of the United States (destCountry field, not included) with the **highest average** passengers count? Show result as {"avgPassengers" : 2312.380, "city" : "Minsk, Belarus" }

**Query**:

db.airlines.aggregate([{

$match: {destCountry: {$not: {$eq: "United States"}}}},

{$group: {\_id: "$destCity", avgPassengers: {$avg: "$passengers"}}},

{$sort: {avgPassengers: -1}},

{$project: {\_id: 0, city: "$\_id", avgPassengers: 1}},

{$limit: 3}

])

**Result**:

{ "avgPassengers" : 8052.380952380952, "city" : "Abu Dhabi, United Arab Emirates" }

{ "avgPassengers" : 7176.596638655462, "city" : "Dubai, United Arab Emirates" }

{ "avgPassengers" : 7103.333333333333, "city" : "Guangzhou, China" }

3. Which carriers provide flights to Latvia (destCountry)? Show result as one document { "\_id" : "Latvia", "carriers" : [ "carrier1", " carrier2", …] }

**Query**:

db.airlines.aggregate([

{$match: {destCountry: "Latvia"}},

{$group: {

\_id: "$destCountry",

carriers: {$addToSet: "$carrier"}

}}

])

**Result:**

{ "\_id" : "Latvia", "carriers" : [ "Uzbekistan Airways", "Blue Jet SP Z o o", "JetClub AG" ] }

4. What are the carriers which flue the most number of passengers from the United State to either Greece, Italy or Spain? Find top 10 carriers, but provide the last 7 carriers (do not include the first 3). Show result as { "\_id" : "<carrier>", "total" : 999}

**Query**:

db.airlines.aggregate([

{$match: {

originCountry: {$eq: "United States"},

destCountry: {$in : ["Greece", "Italy", "Spain"]}

}},

{$group: {

\_id: "$carrier",

total: {$sum: "$passengers"}

}},

{$sort: {total : -1}},

{$limit: 10},

{$skip: 3}

])

**Result**:

{ "\_id" : "Compagnia Aerea Italiana", "total" : 280256 }

{ "\_id" : "United Air Lines Inc.", "total" : 229936 }

{ "\_id" : "Emirates", "total" : 100903 }

{ "\_id" : "Air Europa", "total" : 94968 }

{ "\_id" : "Meridiana S.p.A", "total" : 20308 }

{ "\_id" : "Norwegian Air Shuttle ASA", "total" : 13344 }

{ "\_id" : "VistaJet Limited", "total" : 183 }

5. Find the city (originCity) with the highest sum of passengers for each state (originState) of the United States (originCountry). Provide the city for the first 5 states ordered by state alphabetically (you should see the city for Alaska, Arizona and etc). Show result as { "totalPassengers" : 999, "location" : { "state" : "abc", "city" : "xyz" } }

**Query:**

db.airlines.aggregate([

{$match: {originCountry: {$eq: "United States"}}},

{$group:

{ \_id: {state: "$originState", city: "$originCity"},

totalPassengers: {$sum: "$passengers"}}},

{$sort: {"\_id.state": 1, "totalPassengers" : -1}},

{$group: {

\_id: "$\_id.state",

city: {$first: "$\_id.city"},

totalPassengers: {$first: "$totalPassengers"}

}},

{$sort: {\_id: 1}},

{$project: {\_id: 0, location: {state: "$\_id", city: "$city"}, totalPassengers: 1}},

{$limit: 5}

])

**Result**:

{ "totalPassengers" : 760120, "location" : { "state" : "Alabama", "city" : "Birmingham, AL" } }

{ "totalPassengers" : 1472404, "location" : { "state" : "Alaska", "city" : "Anchorage, AK" } }

{ "totalPassengers" : 13152753, "location" : { "state" : "Arizona", "city" : "Phoenix, AZ" } }

{ "totalPassengers" : 571452, "location" : { "state" : "Arkansas", "city" : "Little Rock, AR" } }

{ "totalPassengers" : 23701556, "location" : { "state" : "California", "city" : "Los Angeles, CA" } }

## Aggregate Enron Collection

For this task you will use the aggregation framework to figure out pairs of people that tend to communicate a lot. To do this, you will need to unwind the To list for each message.

This problem is a little tricky because a recipient may appear more than once in the To list for a message. You will need to fix that in a stage of the aggregation before doing your grouping and counting of (sender, recipient) pairs.

Which pair of people have the greatest number of messages in the dataset?

**Query:**

db.enron.aggregate([

{$unwind: "$headers.To"},

{$group: {

\_id: {

messageId: "$headers.Message-ID",

sender: "$headers.From",

recipient: "$headers.To"

}

}},

{$group: {

\_id:

{sender: "$\_id.sender",

recipient: "$\_id.recipient"},

messagesAmount: {$sum: 1}

}},

{$sort: {messagesAmount: -1}},

{$limit: 1},

{$project: {

\_id: 0,

sender: "$\_id.sender",

recipient: "$\_id.recipient",

messagesAmount: 1

}}],

{allowDiskUse: true

})

**Result:**

{ "messagesAmount" : 750, "sender" : "susan.mara@enron.com", "recipient" : "jeff.dasovich@enron.com" }