#Change Row 45 location

# - entries with # indicate name of variable in database

**Initial data loading**

1. From DB1B – used Q1 Coupon Database
   1. Filtered and included only US origin and destination entries.
2. From DB1B – used Q1 Ticket Database
   1. Filtered and included only US origin (No destination column in Ticket).
3. Removed some variables from both databases.
4. Combined the two databases.

**Data Cleaning**

1. Filtered the database to only include fare (#ItinFare) => 20 (based on Borenstein)
2. Filtered the database to only include fare (#ItinFare) <= 9998 (based on Borenstein)
3. Removed all entries indicating Bulk Fare values (#BulkFare = 1)
4. Removed all entries indicating (#DollarCred = 0)
5. Removed all entries with coupon greater than 4 (based on Borenstein)
6. Removed all entries for one way trip with coupons greater than 2 (based on Borenstein)
7. Cleaning on #FareClass
   1. Changed names.
   2. Removed entries with no fareclass and #FareClass = “U” (no explanation given for U in database)
   3. Combined restricted and unrestricted fare class entries. Classes left now are Economy, Premium and First
8. Converted fare values from per mile to per km (#FarePerMile to #FarePerKM)
9. Removed First Class and changed some column names.

**Combinations**

1. Grouped entries by the combination ***Origin\_Destination\_OperatingCarrier\_FareClass*** and summarized by ***Total Passengers***.
2. Removed all combinations with passengers less than 100.
3. Also removed all combinations which have only 1 #FareClass.
4. Also removed all entries from the non-combined database [@Data Cleaning -9] based on the above two points.

**Percentiles and Normalizing**

1. Calculated percentiles (1000) for #FarePerKM for each combination of-
   1. ***Origin\_Destination\_OperatingCarrier\_Economy***
   2. ***Origin\_Destination\_OperatingCarrier\_Premium***
   3. ***Origin\_Destination\_OperatingCarrier\_All***
2. Repeated the above for mean #FarePerKM for the above 3 combinations.
3. Combined the above two databases.
4. Normalized each percentile of ***Origin\_Destination\_OperatingCarrier*** by dividing by mean #FarePerKM for the combination (illustrated below).
5. Final output in CSV file (q1\_Final\_June2021.csv)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Non- Normalized | | | | | | |
|  | NYC\_LAX\_AA\_E | NYC\_LAX\_AA\_P | NYC\_LAX\_AA\_ALL | DNW\_CHG\_DL\_E | DNW\_CHG\_DL\_P | DNW\_CHG\_DL\_ALL |
| mean | 100 | 200 | 999 | 200 | 300 | 888 |
| Percentile | 10 | 12 | 20 | 20 | 18 | 30 |
| Percentile | 15 | 18 | 30 | 25 | 20 | 40 |
|  |  |  |  |  |  |  |
| NORMALIZED | | | | | | |
|  | NYC\_LAX\_AA\_E | NYC\_LAX\_AA\_P | NYC\_LAX\_AA\_ALL | DNW\_CHG\_DL\_E | DNW\_CHG\_DL\_P | DNW\_CHG\_DL\_ALL |
| Percentile | = 10/999 | = 12/999 | = 20/999 | = 20/888 | = 18/888 | = 30/888 |
| Percentile | = 15/999 | = 18/999 | = 30/999 | = 25/888 | = 20/888 | = 40/888 |

**Comparing with Severin Borenstein** (<http://faculty.haas.berkeley.edu/borenste/airdata.html>)

**Criteria for inclusion of tickets:**

* Tickets with an international segment are excluded – **Done.**
* First-class tickets are excluded for carriers that report less than 90% of tickets as first class (retains FC for carriers that report all tickets as FC in that quarter)- **All First Class excluded.**
* Tickets must be one-way or round-trip; open-jaw, circle trips, etc are excluded – **Could not find information about open-jaw, circle trips**.
* A ticket must have no more than 2 coupons for a one-way trip, no more than 4 coupons (*and no more than 2 coupons each way*) for a rountd-trip ticket – **Could not find a way to carry out the criteria in red**. **Other two criteria are included**.
* Tickets with fare less than $20 or fares above $9998 excluded – **Done** (*Also found some values above 9998 – approx. 100-200 per quarter. Somehow had missed those in earlier versions*).
* Tickets with fares more than 5 times USDOT's Standard Industry Fare Level for observed trip distance during observed quarter are excluded – **Not sure if this was needed. Also, could not find the standard fare value database.**
* Records are for one-way trips, so round-trip tickets are split into two one-way observations. -**Done (as we are doing origin\_destination pair)**