|  |
| --- |
| **Air Traffic Benchmark**  **Specification** |
| Ying Zhang  Niels Nes  Sjoerd Mullender  Martin Kersten |

©2013-2019 MonetDB Solutions. All Rights Reserved.

Copyright in any or all of this documentation belongs to MonetDB Solutions, A. Hepburnstr 83, 1325NS, Almere, the Netherlands (the 'Owner') and may not be used, sold, transferred, copied or reproduced in whole or in part, in any manner of form or on any media to any person other than in accordance with the terms of the Owner's agreement or otherwise without the prior written consent of the Owner.

Table of Contents

Table of Contents 3

1 Introduction 4

1.1 Data Set 4

1.2 Benchmark Queries 5

References 8

Appendix I Air Traffic Benchmark Queries 9

Appendix II Statistics of SF1 Data Set 17

Appendix III Air Traffic Data Schema 20

1. Introduction

The air traffic benchmark uses the U.S. air traffic data, which consist of flight arrival and departure details of all commercial flights reported each month to the U.S. Department of Transportation and Bureau of Transportation Statistics [[1]](#footnote-1) by all U.S. air carriers.

The air traffic data have gained increasing interests in the data analysis and statistics communities in recent years ([1-7]). Compared with the TPC-H benchmark, the air traffic data have several advantages:

1. It is a real-world data set. It has been well known that synthetic data sometimes have little in common with real data ([8]). Therefore, conclusions drawn from benchmarks using synthetic data might not actually translate to expected behaviours in real-world settings.
2. The air traffic data can be scaled up naturally and realistically, because new data are added every month, and this data set can be easily combined with other data sets, such as weather and fuel consumption data.
3. The air traffic data have high practical use, but have not been as thoroughly analysed in the database community as TPC-H. Thus, the chance is considerable that we can gain new insights by analysing the air traffic data.
   1. Data Set

|  |  |
| --- | --- |
| #Flights | 146,542,414 |
| #Delayed departures | 58,589,558 |
| #Delayed arrivals | 66,004,034 |
| #Cancelled flights | 2,571,801 |
| #Diverted flights | 339,856 |
| #Flight dates | 9,071 |
| #Carriers | 32 |
| #Airplanes | 10,553 |
| #Airports | 380 |
| #Cities | 372 |
| Average distance between airport (miles) | 712 |

Table 1: statistics of the basic data set

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scale Factor** | **#CSV files** | **CSV size (GB)** | **DB size (GB)** | **#tuples** |
| *SF1* | 301 | 60 | 37 | 146,542,414 |
| *SF2* | 602 | 120 | 71 | 293,084,828 |

Table 2: statistics of different scale factors

The air traffic data in CSV format can be downloaded from [here](http://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time)[[2]](#footnote-2). The data of each month are stored in a separate CSV file. For this project, we have downloaded the data from January 1988 to March 2014. This results in 318 CSV files, with a total size of ~60 GB on disk. However, because 17 files contain errors, e.g., invalid characters, only 301 files are used in the project. Table 1 lists some statistics of the basic data set of 301 files, where “DB size” is the size of the binary equivalent of the CSV files on disk, after the CSV data have been loaded into MonetDB. More statistics of the data is given in Appendix II.

***Scale Factors.*** We call the basic data set containing the 301 files downloaded from the BTS website the SF1 data set. Data are duplicated to create larger data sets. Table 2 shows the sizes the data sets of different scale factors.

***Data Schema.*** All data of one data set are store in a single table in the database, called *ontime*, containing 109 columns. The complete schema of this table is attached in Appendix IV.

* 1. Benchmark Queries

The benchmark contains a set of 19 queries based on popular statistic analysis use cases that have been applied on the air traffic data before (see e.g., [1-7]). The existing queries and use cases are first merged with duplicates removed. Then, 19 queries ([9]) are carefully selected such that they both contain interesting real-life use cases and address a wide variety of DBMS features.

Below we give a briefly description of each query and its properties. The SQL implementations of all queries are included in Appendix II. Table 3 gives an overview of the main SQL features addressed by each query, the number of tuples each query reads from the input table *ontime* (“#tuples input”), and the number of tuples each query returns as its output (“#tuples output”). The tuple counts are computed based on the data set SF1.

1. *Whole data set statistics*

Computes the number of unique airlines and airports, the number of cancelled flights and flights with long delays (i.e., longer than 15 minutes).

1. *Summarise arrival delays*

This query mimics the R summary() function by computing the min, max, median, mean, quantile 0.05 and quantile 0.5 of the arrival delays

1. *Distribution of flights per month*

Computes the number of flights for each month in the data set.

1. *Distribution of long delays by day of week*

This query computes the number of flights with long delays (i.e., longer than15 minutes) for each day of week.

1. *Proportion of long delays per year*

For each year, this query computes the percentage of flights with long delays (i.e., longer than 15 minutes) against all flights.

1. *Top 10 best-connected cities*

For each destination city, compute the number of unique origin cities. Return the ten destination cities with the highest count.

1. *Histogram of arrival delays.*

Divide the value of the arrival delays into 30 value categories, and compute the count for each group.

1. *Compare arrival delays of weekdays vs. weekends.*

For every departure hour, compute its average arrival delay and the lower and upper bounds of confidence. Compare departing hours during weekdays and weekends.

1. *Detect cascading delays from one airport to another*

For all flights within an arbitrary day, find chains of three airports, where airport2 is a destination of airport1 and airport3 is a destination of airport2, and they all have some long departure/arrival delays (i.e., langer than 15 minutes). Return a chain of airports, if there is a flight whose arrival time at airport2 is smaller than a flight’s departure time from airport2. This indicates that the delay from airport1 to airport2 might have cause the delay from airport2 to airport3.

1. *Compare flight patterns of all flights to and from a major airport (e.g. Chicago ORD).*

For each airport connected to the chosen airport, compute the number of inbound and/or outbound flights.

1. *Compare flight patterns before and after 9/11 (2001)*

For every two connected airports, compute the total number of flights (inbound + outbound flights) between these airports one year before vs. after 9/11.

1. *Proportion of long delays and cancellations per day*

For every flight date, compute the percentages of flight departures delayed longer than 15 minutes and cancelled flights.

1. *Proportion of long departure delays of a major airport in a year.*

Given an airport (e.g., ORD), compute the percentage of departures delayed longer than 15 minutes to each unique destination in the given year (e.g., 2013).

1. *Yearly median long delays per airport.*

For each airport and each year, compute the median delay of flights delayed at least 15 minutes for both inbound (arriving) and outbound (departing) flights.

1. *Predict carrier delay profiles*

For each carrier, for each scheduled departure hour of the given day (e.g., October 24, 2007), compute the mean of arrival delays.

1. *Daily flight volumes of major airports (e.g., Atlanta ATL, Chicago ORD, Seattle SEA, St Louis STL).*

For each of the major airport, for each day of a year, compute the average of the total number of flights per day.

1. *Predict arrival delays per airport with confidence.*

Compute the mean and standard error of arrival delays for each destination Airport (and compute the standard error)

1. *Changes of carrier popularity, in terms of volume, over time*

For each carrier, compute the total number of flights per year carried out by this carrier.

1. *Logistic regression of flight delay*

Compute the probability that a flight will be delayed for more than 15 minutes, given its departing time, departing and arriving airports, and carrier.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **GROUP BY** | **ORDER BY** | **Aggregation functions** | **JOIN** | **UNION** | **LIMIT** | **#tuples input** | **#tuples output** |
| *q1* |  |  | COUNT, SUM |  |  |  | 146542414 | 1 |
| *q2* |  |  | MIN, MAX, AVG, QUANTILE |  |  |  | 146542414 | 1 |
| *q3* | ✓ | ✓ | COUNT |  |  |  | 146542414 | 298 |
| *q4* | ✓ | ✓ | COUNT |  |  |  | 22645828 | 7 |
| *q5* | ✓ | ✓ | COUNT | EQUAL JOIN |  |  | 146542414 | 26 |
| *q6* | ✓ | ✓ | COUNT |  |  | ✓ | 146542414 | 10 |
| *q7* | ✓ | ✓ | MIN, MAX, SUM |  |  |  | 143630757 | 30 |
| *q8* | ✓ | ✓ | AVG, STDDEV\_SAMP |  | ✓ |  | 146542414 | 49 |
| *q9* | ✓ | ✓ | AVG | EQUAL JOIN |  |  | 4309 | 56312 |
| *q10* | ✓ |  | COUNT | FULL OUTER JOIN |  |  | 15391774 | 195 |
| *q11* | ✓ |  | COUNT | FULL OUTER JOIN |  |  | 12295086 | 2550 |
| *q12* | ✓ | ✓ | COUNT, SUM |  |  |  | 146542414 | 9071 |
| *q13* | ✓ | ✓ | COUNT, SUM |  |  |  | 306439 | 150 |
| *q14* | ✓ | ✓ | MEDIAN | FULL OUTER JOIN |  |  | 22645828 | 6671 |
| *q15* | ✓ | ✓ | AVG |  | ✓ |  | 20903 | 480 |
| *q16* | ✓ | ✓ | AVG, SUM, COUNT | LEFT JOIN | ✓ |  | 39873185 | 1464 |
| *q17* | ✓ | ✓ | AVG, STDDEV\_SAMP |  |  |  | 143630757 | 373 |
| *q18* | ✓ |  | COUNT |  |  |  | 146542414 | 369 |
| *q19* | ✓ | ✓ | COUNT, SUM |  |  |  | 146542414 | 199956 |

Table 3: main SQL features used by the benchmark queries; the number of tuples each query reads from the data table *ontime* (“#tuples input”); and the number of tuples each query returns (“#tuples output”). The tuple counts are computed based on the SF1 data set.

References

1. *A simple Big Data analysis using the RevoScaleR package in Revolution R*. Stephen Weller, March 24, 2011. <http://blog.revolutionanalytics.com/2011/05/big-data-analysis-in-revolution-r.html>
2. *Airline Dataset, RHIPE 0.65.2 documentation*. Saptarshi Guha, 2010. <http://www.stat.purdue.edu/~sguha/rhipe/doc/html/airline.html#id2>
3. *Airline on-time performance* challenge. The American Statistical Association (ASA) Sections on Statistical Computing Statistical Graphics, 2009. <http://stat-computing.org/dataexpo/2009/>, <http://www.amstat.org/>
4. *Congestion in the sky: Visualising domestic airline traffic with SAS*. Rick Wicklin and Robert Allison, SAS Institute. <http://stat-computing.org/dataexpo/2009/posters/wicklin-allison.pdf>
5. *Delayed, Cancelled, On-Time, Boarding ... Flying in the USA*. Heike Hofmann, Di Cook, Chris Kielion, Barret Schloerke, Jon Hobbs, Adam Loy, Lawrence Mosley, David Rockoff, Yuanyuan Huang, Danielle Wrolstad and Tengfei Yin, Iowa State University. <http://stat-computing.org/dataexpo/2009/posters/hofmann-cook.pdf>
6. *Quick comparison of MyISAM, Infobright, and MonetDB*. Baron Schwartz, September 29, 2009. <http://www.mysqlperformanceblog.com/2009/09/29/quick-comparison-of-myisam-infobright-and-monetdb/>
7. *Analyzing air traffic performance with InfoBright and MonetDB*. Vadim Tkachenko, October 2, 2009. <http://www.mysqlperformanceblog.com/2009/10/02/analyzing-air-traffic-performance-with-infobright-and-monetdb/>
8. S. Duan, A. Kementsietsidis, K. Srinivas, and O. Udrea. *Apples and Oranges: A Comparison  of RDF Benchmarks and Real RDF Datasets*. In *SIGMOD*, 2011.
9. Alexander S. Szalay, José A. Blakeley. *Gray’s Laws: Database-centric Computing in Science*. In The Fourth Paradigm: Data-Intensive Scientific  Discovery. T. Hey, S. Tansley, and K. Tolle, editors. Microsoft Research, Oct. 2009.

1. Air Traffic Benchmark Queries

-- q1. Overview statistics: #airlines, #Airports, #cancelled-flights, #long-delayed-flights

SELECT COUNT(DISTINCT "AirlineID") AS "Airlines",

COUNT(DISTINCT "Origin") AS "Airports",

CAST(SUM("Cancelled") AS BIGINT) AS "CancelledFlights",

CAST(SUM("DepDel15") AS BIGINT) AS "LongDepDelays"

FROM ontime;

-- q2. Summarise the arrival delays (mimics the R summary() function)

SELECT MIN("ArrDelay") AS "MinArrDelay",

CAST(QUANTILE("ArrDelay", 0.05) AS DECIMAL(8,2)) AS "Q05ArrDelay",

CAST(QUANTILE("ArrDelay", 0.5) AS DECIMAL(8,2)) AS "MedianArrDelay",

CAST(AVG("ArrDelay") AS DECIMAL(8,2)) AS "MeanArrDelay",

CAST(QUANTILE("ArrDelay", 0.95) AS DECIMAL(8,2)) AS "Q95ArrDelay",

MAX("ArrDelay") AS "MaxArrDelay"

FROM ontime;

-- q3. Distribution of flights per month

SELECT "Year", "Month", COUNT(\*) AS "Flights"

FROM ontime

GROUP BY "Year", "Month"

ORDER BY "Year", "Month";

-- q4. Distribution of long delays (> 15min) by day of week

SELECT "DayOfWeek", COUNT(\*) AS "Flights"

FROM ontime

WHERE "DepDelay" > 15

GROUP BY "DayOfWeek"

ORDER BY "DayOfWeek";

-- q5. Proportion of long delays per airport per year (or more precise: % inbound delays per year and Airports with most/least delays?)

WITH t AS ( -- #long\_delays/year

SELECT "Year", COUNT(\*) AS c1

FROM ontime

WHERE "DepDelay" > 15

GROUP BY "Year"

),

t2 AS (-- #delays/year

SELECT "Year", COUNT(\*) AS c2

FROM ontime

GROUP BY "Year"

)

SELECT t."Year", CAST(c1 AS DECIMAL(16,2))/c2\*100 AS "ProportionLongDelay"

FROM t JOIN t2 ON (t."Year" = t2."Year")

ORDER BY t."Year";

-- q6. Top 10 best-connected cities

SELECT "DestCityName" AS "City", COUNT(DISTINCT "OriginCityName") AS "Origins"

FROM ontime

GROUP BY "DestCityName"

ORDER BY "Origins" DESC

LIMIT 10;

-- q7. Compute a histogram of arrival delay, which values are divided into "categories" number of groups.

CREATE FUNCTION histo (categories TINYINT)

RETURNS TABLE (low INT, high INT, CntArrDelay INT)

BEGIN

DECLARE minAD INT, sz INT, grp INT;

SET minAD = (SELECT min("ArrDelay") FROM ontime);

SET sz = (SELECT (max("ArrDelay") - minAD)/categories FROM ontime);

SET grp = 0;

DECLARE TABLE tmp1(low INT, "CntArrDelay" INT DEFAULT 0);

WHILE (grp < categories) DO

INSERT INTO tmp1(low) VALUES (minAD + sz \* grp);

SET grp = grp + 1;

END WHILE;

-- devide the ArrDelay values into groups

INSERT INTO tmp1

SELECT low, COUNT(\*) AS "CntArrDelay"

FROM (

SELECT minAD + sz \* CAST(FLOOR(("ArrDelay" - minAD) / sz) AS INT) AS low

FROM ontime

WHERE "ArrDelay" IS NOT NULL) AS t

GROUP BY low;

RETURN

SELECT low, low + sz AS high, SUM("CntArrDelay") AS "CntArrDelay"

FROM tmp1

GROUP BY low

ORDER BY low;

END; -- FUNCTION histo()

SELECT \* FROM histo(30);

-- q8. Arrival Delay by Departure Hour: Weekdays and Weekends .

WITH t1 AS(

SELECT 'Weekend' AS "DayOfWeek",

CAST (FLOOR("CRSDepTime"%2400/100) AS INT) AS "Hour",

CAST(AVG("ArrDelay") AS DECIMAL(8,2)) AS "ArrDelay",

CAST(AVG("ArrDelay") - 2 \* STDDEV\_SAMP("ArrDelay") AS DECIMAL(8,2))

AS "lowerConfBound",

CAST(AVG("ArrDelay") + 2 \* STDDEV\_SAMP("ArrDelay") AS DECIMAL(8,2))

AS "upperConfBound"

FROM ontime

WHERE "DayOfWeek" in (6, 7) -- Saturday or Sunday

GROUP BY "Hour"

),

t2 AS(

SELECT 'Weekday' AS "DayOfWeek",

CAST (FLOOR("CRSDepTime"%2400/100) AS INT) AS "Hour",

CAST(AVG("ArrDelay") AS DECIMAL(8,2)) AS "ArrDelay",

CAST(AVG("ArrDelay") - 2 \* STDDEV\_SAMP("ArrDelay") AS DECIMAL(8,2))

AS "lowerConfBound",

CAST(AVG("ArrDelay") + 2 \* STDDEV\_SAMP("ArrDelay") AS DECIMAL(8,2))

AS "upperConfBound"

FROM ontime

WHERE "DayOfWeek" in (1,2,3,4,5) -- Monday to Friday

GROUP BY "Hour"

)

SELECT \* FROM t1 UNION ALL SELECT \* FROM t2

ORDER BY "DayOfWeek", "Hour";

-- q9. Can you detect cascading failures as delays in one Airport create delays in others? Are there critical links in the system?”

WITH t1 AS (

SELECT "Origin", "CRSDepTime", "DepDelay", "Dest", "ArrDelay", "CRSArrTime"

FROM ontime

WHERE "DepDelay" > 15 AND "ArrDelay" > 15

-- we only look at the data of one day

AND "Month" = 3 AND "DayofMonth" = 24 AND "Year" = 2013

)

SELECT t1."Origin" AS "Airport1",

CAST(AVG(t1."DepDelay") AS DECIMAL(8,2)) AS "AVGDepDelay",

CAST(AVG(t1."ArrDelay") AS DECIMAL(8,2)) AS "AVGArrDelay",

t2."Origin" AS "Airport2",

CAST(AVG(t2."DepDelay") AS DECIMAL(8,2)) AS "AVGDepDelay2",

CAST(AVG(t2."ArrDelay") AS DECIMAL(8,2)) AS "AVGArrDelay2",

t2."Dest" AS "Airport3"

FROM t1, t1 AS t2

WHERE t1."Dest" = t2."Origin" AND t1."CRSArrTime" < t2."CRSDepTime"

GROUP BY t1."Origin", t2."Origin", t2."Dest"

ORDER BY "AVGDepDelay" DESC, "AVGArrDelay" DESC,

"AVGDepDelay2" DESC, "AVGArrDelay2" DESC;

-- q10. Compare flight patterns of all flights to and from a major Airport like Chicago (ORD)

WITH t1 AS ( -- flights to ORD

SELECT "Origin" AS ap, COUNT(\*) AS cnt\_in

FROM ontime

WHERE "Dest" = 'ORD'

GROUP BY "Origin"

),

t2 AS ( -- flights from ORD

SELECT "Dest" AS ap, COUNT(\*) AS cnt\_out

FROM ontime

WHERE "Origin" = 'ORD'

GROUP BY "Dest"

),

t3 AS ( -- merge t1, t2 into one table

SELECT t1.ap AS ap1, t1.cnt\_in, t2.ap AS ap2, t2.cnt\_out

FROM t1 FULL OUTER JOIN t2 ON (t1.ap = t2.ap))

SELECT CASE WHEN ap1 IS NULL THEN ap2 ELSE ap1 END AS "Airport",

cnt\_in AS "InboundFlights", cnt\_out AS "OutboundFlights"

FROM t3;

-- q11. Compare flight patterns 1 year before and after 9/11 (2001)

WITH t1 AS ( -- #flights per route before 9/11

SELECT SQL\_MIN("Origin", "Dest") || ' <-> ' ||

SQL\_MAX("Origin", "Dest") AS route,

COUNT(\*) AS cnt\_before

FROM ontime

WHERE '2010-09-11' < "FlightDate" AND "FlightDate" < '2011-09-11'

GROUP BY route

),

t2 AS ( -- #flights per route after 9/11

SELECT SQL\_MIN("Origin", "Dest") || ' <-> ' ||

SQL\_MAX("Origin", "Dest") AS route,

COUNT(\*) AS cnt\_after

FROM ontime

WHERE '2011-09-11' <= "FlightDate" AND "FlightDate" < '2012-09-11'

GROUP BY route

),

t3 AS ( -- merge t1, t2 into one table

SELECT t1.route AS route1, t1.cnt\_before, t2.route AS route2, t2.cnt\_after

FROM t1 FULL OUTER JOIN t2 ON (t1.route = t2.route)

)

SELECT CASE WHEN route1 IS NULL THEN route2 ELSE route1 END AS "Route",

cnt\_before AS "FlightsBefore", cnt\_after AS "FlightsAfter"

FROM t3;

-- q12. % of flight departures delayed > 15 min, and % of flights cancelled per day

SELECT "FlightDate",

CAST(SUM("DepDel15") AS DECIMAL(8,2))/COUNT(\*)\*100 AS "PercLongDelay",

CAST(SUM("Cancelled") AS DECIMAL(8,2))/COUNT(\*)\*100 AS "PercCancelled"

FROM ontime

GROUP BY "FlightDate"

ORDER BY "FlightDate";

-- q13. Given an Airport (e.g., ORD), % delayed departures > 15 min to each destination in 2013

SELECT "Dest", CAST(SUM("DepDel15") AS DECIMAL(16,2))/COUNT(\*)\*100

AS "ProportionLongDelay"

FROM ontime

WHERE "Origin" = 'ORD' AND "Year" = 2013

GROUP BY "Dest"

ORDER BY "ProportionLongDelay" DESC;

-- q14. For each Airport and each year, compute the median delay of flights delayed at least 15 minutes for both inbound (arriving) and outbound (departing) flights

WITH t1 AS ( -- median of outbound delays

SELECT "Year", "Origin" AS "Airport",

MEDIAN("DepDelay") AS "MedianOutboundDelay"

FROM ontime

WHERE "DepDelay" > 15

GROUP BY "Year", "Origin"

),

t2 AS ( -- median of inbound delays

SELECT "Year", "Dest" AS "Airport",

MEDIAN("ArrDelay") AS "MedianInboundDelay"

FROM ontime

WHERE "DepDelay" > 15

GROUP BY "Year", "Dest"

),

t3 AS ( -- merge t1, t2 into one table

SELECT t1."Year" AS y1, t1."Airport" AS a1, "MedianOutboundDelay",

t2."Year" AS y2, t2."Airport" AS a2, "MedianInboundDelay"

FROM t1 FULL OUTER JOIN t2 ON

(t1."Year" = t2. "Year" AND t1."Airport" = t2."Airport")

)

SELECT CASE WHEN y1 IS NULL THEN y2 ELSE y1 END AS "Year",

CASE WHEN a1 IS NULL THEN a2 ELSE a1 END AS "Airport",

"MedianOutboundDelay", "MedianInboundDelay"

FROM t3

ORDER BY "Year", "Airport";

-- q15. Carrier delay profiles: for each carrier, on 24oct2007, for each scheduled departure time, compute the predicted arrival delay.

CREATE TABLE tmp (

"Hour" TINYINT,

"PredictedArrDelay" DECIMAL(8,2) DEFAULT 0.0

);

INSERT INTO tmp("Hour") VALUES

(0), (1), (2), (3), (4), (5), (6), (7), (8), (9),

(10), (11), (12), (13), (14), (15), (16), (17), (18), (19),

(20), (21), (22), (23);

WITH t1 AS (

SELECT "Carrier", CAST (FLOOR("CRSDepTime"%2400/100) AS INT) AS "Hour",

CAST(AVG("ArrDelay") AS DECIMAL(8,2)) AS "PredictedArrDelay"

FROM ontime

WHERE "Year" = 2007 AND "Month" = 10 AND "DayofMonth" = 24

GROUP BY "Carrier", "Hour"

),

t2 AS (

SELECT t."Carrier", tmp.\*

FROM tmp, (SELECT DISTINCT "Carrier" FROM t1) AS t

)

SELECT "Carrier", "Hour", SUM("PredictedArrDelay")

FROM (SELECT \* FROM t1 UNION SELECT \* FROM t2) AS t

GROUP BY "Carrier", "Hour"

ORDER BY "Carrier", "Hour";

-- q16. Volume at Major Airports: Atlanta, Chicago, Seattle, St Louis: average #flights per Airport per day over all years

WITH t1 AS ( -- outbound flights of the Major Airports

SELECT "Origin" AS "Airport", "Month", "DayofMonth", COUNT(\*) AS "Count"

FROM ontime

WHERE "Origin" = 'ATL' OR "Origin" = 'ORD' OR

"Origin" = 'SEA' OR "Origin" = 'STL'

GROUP BY "Origin", "Month", "DayofMonth"

),

t2 AS ( -- inbound flights of the Major Airports

SELECT "Dest" AS "Airport", "Month", "DayofMonth", COUNT(\*) AS "Count"

FROM ontime

WHERE "Dest" = 'ATL' OR "Dest" = 'ORD' OR "Dest" = 'SEA' OR "Dest" = 'STL'

GROUP BY "Dest", "Month", "DayofMonth"

),

t3 AS ( -- since each day doesn't appear the same number of times in the database

SELECT "Month", "DayofMonth", COUNT(\*) AS cnt

FROM (SELECT DISTINCT "Year", "Month", "DayofMonth" FROM ontime) AS t

GROUP BY "Month", "DayofMonth"

)

SELECT "Airport", "Month", "DayofMonth",

CAST(SUM(t."Count")/AVG(t3.cnt) AS DECIMAL(8,2)) AS "AvgFlightsPerDay"

-- putting in/out-bound flights together

FROM (SELECT \* FROM t1 UNION ALL SELECT \* FROM t2) AS t LEFT JOIN t3 ON

(t."Month" = t3."Month" AND t."DayofMonth" = t3."DayofMonth")

GROUP BY "Airport", t."Month", t."DayofMonth"

ORDER BY "Airport", t."Month", t."DayofMonth";

-- q17. AVG delay by destination Airport (and compute the standard error):

SELECT "Dest",

CAST(AVG("ArrDelay") AS DECIMAL(8,2)) AS "AvgDelay",

CAST(STDDEV\_SAMP("ArrDelay") AS DECIMAL(8,2)) AS "AvgDelayErr"

FROM ontime

WHERE "ArrDelay" IS NOT NULL

GROUP BY "Dest"

ORDER BY "AvgDelay";

-- q18. Changes of carrier popularity, in terms of volume, over time

SELECT "Carrier", "Year", COUNT(\*)

FROM ontime

GROUP BY "Carrier", "Year";

-- q19. Logistic regression: predict (the probability) if a flight will be delayed (>15 min) based on departing time, departing/arriving Airport, and/or carrier.

SELECT CAST (FLOOR("CRSDepTime"%2400/100) AS INT) AS "Hour",

"Origin", "Dest", "Carrier",

CAST(SUM("DepDel15") AS DOUBLE)/COUNT(\*) >= 0.5 AS "PossibleLongDelay"

FROM ontime

GROUP BY "Origin", "Dest", "Carrier", "Hour"

ORDER BY "PossibleLongDelay", "Hour", "Origin", "Dest", "Carrier";

1. Statistics of SF1 Data Set

+----------------------+----------+-------+-----------+--------+-----------+---------+---------+--------+

| column | type | width | count | unique | nils | minval | maxval | sorted |

+======================+==========+=======+===========+========+===========+=========+=========+========+

| Year | smallint | 2 | 146542414 | 26 | 0 | 1988 | 2014 | true |

| Quarter | tinyint | 1 | 146542414 | 4 | 0 | nil | nil | false |

| Month | tinyint | 1 | 146542414 | 12 | 0 | nil | nil | false |

| DayofMonth | tinyint | 1 | 146542414 | 31 | 0 | nil | nil | false |

| DayOfWeek | tinyint | 1 | 146542414 | 7 | 0 | nil | nil | false |

| FlightDate | date | 4 | 146542414 | 9071 | 0 | nil | nil | false |

| UniqueCarrier | char | 2 | 146542414 | 30 | 0 | nil | nil | false |

| AirlineID | decimal | 4 | 146542414 | 30 | 0 | 1938600 | 2117100 | false |

| Carrier | char | 2 | 146542414 | 32 | 0 | nil | nil | false |

| TailNum | varchar | 4 | 146542414 | 10554 | 0 | nil | nil | false |

| FlightNum | varchar | 3 | 146542414 | 8217 | 0 | nil | nil | false |

| OriginAirportID | varchar | 5 | 146542414 | 377 | 0 | nil | nil | false |

| OriginAirportSeqID | varchar | 7 | 146542414 | 786 | 0 | nil | nil | false |

| OriginCityMarketID | varchar | 5 | 146542414 | 345 | 0 | nil | nil | false |

| Origin | char | 3 | 146542414 | 376 | 0 | nil | nil | false |

| OriginCityName | varchar | 13 | 146542414 | 371 | 0 | nil | nil | false |

| OriginState | char | 1 | 146542414 | 54 | 0 | nil | nil | false |

| OriginStateFips | varchar | 1 | 146542414 | 54 | 0 | nil | nil | false |

| OriginStateName | varchar | 8 | 146542414 | 54 | 0 | nil | nil | false |

| OriginWac | decimal | 4 | 146542414 | 59 | 0 | 100 | 84100 | false |

| DestAirportID | varchar | 5 | 146542414 | 381 | 0 | nil | nil | false |

| DestAirportSeqID | varchar | 7 | 146542414 | 791 | 0 | nil | nil | false |

| DestCityMarketID | varchar | 5 | 146542414 | 346 | 0 | nil | nil | false |

| Dest | char | 3 | 146542414 | 380 | 0 | nil | nil | false |

| DestCityName | varchar | 13 | 146542414 | 372 | 0 | nil | nil | false |

| DestState | char | 1 | 146542414 | 54 | 0 | nil | nil | false |

| DestStateFips | varchar | 1 | 146542414 | 54 | 0 | nil | nil | false |

| DestStateName | varchar | 8 | 146542414 | 54 | 0 | nil | nil | false |

| DestWac | decimal | 4 | 146542414 | 59 | 0 | 100 | 84100 | false |

| CRSDepTime | decimal | 4 | 146542414 | 1405 | 1 | 0 | 240000 | false |

| DepTime | decimal | 4 | 146542414 | 1444 | 2548894 | 100 | 240000 | false |

| DepDelay | decimal | 4 | 146542414 | 1922 | 2548894 | -141000 | 260100 | false |

| DepDelayMinutes | decimal | 4 | 146542414 | 1550 | 2548894 | 0 | 260100 | false |

| DepDel15 | decimal | 4 | 146542414 | 3 | 2548894 | 0 | 100 | false |

| DepartureDelayGroups | decimal | 4 | 146542414 | 16 | 2548894 | -200 | 1200 | false |

| DepTimeBlk | varchar | 8 | 146542414 | 20 | 0 | nil | nil | false |

| TaxiOut | decimal | 4 | 146542414 | 706 | 38099436 | 0 | 390500 | false |

| WheelsOff | decimal | 4 | 146542414 | 1446 | 38102262 | 100 | 723000 | false |

| WheelsOn | decimal | 4 | 146542414 | 1441 | 38297165 | 100 | 240000 | false |

| TaxiIn | decimal | 4 | 146542414 | 1185 | 38288698 | 0 | 144000 | false |

| CRSArrTime | decimal | 4 | 146542414 | 1443 | 2 | 0 | 240000 | false |

| ArrTime | decimal | 4 | 146542414 | 1441 | 2846228 | 100 | 240000 | false |

| ArrDelay | decimal | 4 | 146542414 | 1864 | 2911657 | -143700 | 259800 | false |

| ArrDelayMinutes | decimal | 4 | 146542414 | 1528 | 2911657 | 0 | 259800 | false |

| ArrDel15 | decimal | 4 | 146542414 | 3 | 2911657 | 0 | 100 | false |

| ArrivalDelayGroups | decimal | 4 | 146542414 | 16 | 2911657 | -200 | 1200 | false |

| ArrTimeBlk | varchar | 8 | 146542414 | 20 | 0 | nil | nil | false |

| Cancelled | tinyint | 1 | 146542414 | 2 | 0 | nil | nil | false |

| CancellationCode | char | 0 | 146542414 | 5 | 0 | nil | nil | false |

| Diverted | tinyint | 1 | 146542414 | 2 | 0 | nil | nil | false |

| CRSElapsedTime | decimal | 4 | 146542414 | 767 | 25994 | -16200 | 186500 | false |

| ActualElapsedTime | decimal | 4 | 146542414 | 892 | 2911658 | -71000 | 144000 | false |

| AirTime | decimal | 4 | 146542414 | 1318 | 38354249 | -237800 | 135000 | false |

| Flights | decimal | 4 | 146542414 | 1 | 0 | 100 | 100 | true |

| Distance | decimal | 4 | 146542414 | 1852 | 0 | 0 | 498300 | false |

| DistanceGroup | tinyint | 1 | 146542414 | 11 | 0 | nil | nil | false |

| CarrierDelay | decimal | 4 | 146542414 | 1446 | 132176765 | 0 | 258000 | false |

| WeatherDelay | decimal | 4 | 146542414 | 983 | 132176765 | 0 | 161500 | false |

| NASDelay | decimal | 4 | 146542414 | 900 | 132176765 | -6000 | 139200 | false |

| SecurityDelay | decimal | 4 | 146542414 | 319 | 132176765 | 0 | 78200 | false |

| LateAircraftDelay | decimal | 4 | 146542414 | 922 | 132176765 | 0 | 143700 | false |

| FirstDepTime | varchar | 0 | 146542414 | 1346 | 0 | nil | nil | false |

| TotalAddGTime | varchar | 0 | 146542414 | 312 | 146354141 | nil | nil | false |

| LongestAddGTime | varchar | 0 | 146542414 | 287 | 146354141 | nil | nil | false |

| DivAirportLandings | varchar | 0 | 146542414 | 6 | 112059745 | nil | nil | false |

| DivReachedDest | varchar | 0 | 146542414 | 3 | 146462755 | nil | nil | false |

| DivActualElapsedTime | varchar | 0 | 146542414 | 1193 | 146476984 | nil | nil | false |

| DivArrDelay | varchar | 0 | 146542414 | 1149 | 146476984 | nil | nil | false |

| DivDistance | varchar | 0 | 146542414 | 938 | 146463060 | nil | nil | false |

| Div1Airport | varchar | 0 | 146542414 | 386 | 0 | nil | nil | false |

| Div1AirportID | varchar | 0 | 146542414 | 385 | 146456283 | nil | nil | false |

| Div1AirportSeqID | varchar | 0 | 146542414 | 720 | 146456283 | nil | nil | false |

| Div1WheelsOn | varchar | 0 | 146542414 | 1419 | 0 | nil | nil | false |

| Div1TotalGTime | varchar | 0 | 146542414 | 274 | 146456282 | nil | nil | false |

| Div1LongestGTime | varchar | 0 | 146542414 | 247 | 146456282 | nil | nil | false |

| Div1WheelsOff | varchar | 0 | 146542414 | 1418 | 0 | nil | nil | false |

| Div1TailNum | varchar | 0 | 146542414 | 5742 | 0 | nil | nil | false |

| Div2Airport | varchar | 0 | 146542414 | 141 | 0 | nil | nil | false |

| Div2AirportID | varchar | 0 | 146542414 | 141 | 146541452 | nil | nil | false |

| Div2AirportSeqID | varchar | 0 | 146542414 | 206 | 146541452 | nil | nil | false |

| Div2WheelsOn | varchar | 0 | 146542414 | 634 | 0 | nil | nil | false |

| Div2TotalGTime | varchar | 0 | 146542414 | 96 | 146541452 | nil | nil | false |

| Div2LongestGTime | varchar | 0 | 146542414 | 92 | 146541452 | nil | nil | false |

| Div2WheelsOff | varchar | 0 | 146542414 | 205 | 0 | nil | nil | false |

| Div2TailNum | varchar | 0 | 146542414 | 224 | 0 | nil | nil | false |

| Div3Airport | varchar | 0 | 146542414 | 11 | 0 | nil | nil | false |

| Div3AirportID | varchar | 0 | 146542414 | 11 | 146542404 | nil | nil | false |

| Div3AirportSeqID | varchar | 0 | 146542414 | 11 | 146542404 | nil | nil | false |

| Div3WheelsOn | varchar | 0 | 146542414 | 11 | 0 | nil | nil | false |

| Div3TotalGTime | varchar | 0 | 146542414 | 7 | 146542404 | nil | nil | false |

| Div3LongestGTime | varchar | 0 | 146542414 | 7 | 146542404 | nil | nil | false |

| Div3WheelsOff | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div3TailNum | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div4Airport | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div4AirportID | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div4AirportSeqID | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div4WheelsOn | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div4TotalGTime | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div4LongestGTime | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div4WheelsOff | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div4TailNum | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div5Airport | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div5AirportID | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div5AirportSeqID | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div5WheelsOn | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div5TotalGTime | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div5LongestGTime | varchar | 0 | 146542414 | 1 | 146542414 | nil | nil | true |

| Div5WheelsOff | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

| Div5TailNum | varchar | 0 | 146542414 | 1 | 0 | nil | nil | true |

+----------------------+----------+-------+-----------+--------+-----------+---------+---------+--------+

1. Air Traffic Data Schema

CREATE TABLE ontime(

-- 1 Year

Year SMALLINT DEFAULT NULL,

-- 2 Quarter (1-4)

Quarter TINYINT DEFAULT NULL,

-- 3 Month

Month TINYINT DEFAULT NULL,

-- 4 Day of Month

DayofMonth TINYINT DEFAULT NULL,

-- 5 Day of Week

DayOfWeek TINYINT DEFAULT NULL,

-- 6 Flight Date (yyyymmdd)

FlightDate DATE DEFAULT NULL,

-- 7 Unique Carrier Code.

UniqueCarrier CHAR(7) DEFAULT NULL,

-- field for analysis across a range of years.

-- 8 An identification number assigned by US DOT to identify a unique airline.

AirlineID DECIMAL(8,2) DEFAULT NULL,

-- 9 Code assigned by IATA and commonly used to identify a carrier.

Carrier CHAR(2) DEFAULT NULL,

-- 10 Tail Number

TailNum VARCHAR(50) DEFAULT NULL,

-- 11 Flight Number

FlightNum VARCHAR(10) DEFAULT NULL,

-- 12 Origin Airport, Airport ID.

OriginAirportID VARCHAR (10) DEFAULT NULL,

-- 13 Origin Airport, Airport Sequence ID.

OriginAirportSeqID VARCHAR (10) DEFAULT NULL,

-- 14 Origin Airport, City Market ID.

OriginCityMarketID VARCHAR (10) DEFAULT NULL,

-- 15 Origin Airport

Origin CHAR(5) DEFAULT NULL,

-- 16 Origin Airport, City Name

OriginCityName VARCHAR(100) DEFAULT NULL,

-- 17 Origin Airport, State Code

OriginState CHAR(2) DEFAULT NULL,

-- 18 Origin Airport, State Fips

OriginStateFips VARCHAR (10) DEFAULT NULL,

-- 19 Origin Airport, State Name

OriginStateName VARCHAR(100) DEFAULT NULL,

-- 20 Origin Airport, World Area Code

OriginWac DECIMAL(8,2) DEFAULT NULL,

-- 21 Destination Airport, Airport ID.

DestAirportID VARCHAR (10) DEFAULT NULL,

-- 22 Destination Airport, Airport Sequence ID.

DestAirportSeqID VARCHAR (10) DEFAULT NULL,

-- 23 Destination Airport, City Market ID.

DestCityMarketID VARCHAR (10) DEFAULT NULL,

-- 24 Destination Airport

Dest CHAR(5) DEFAULT NULL,

-- 25 Destination Airport, City Name

DestCityName VARCHAR(100) DEFAULT NULL,

-- 26 Destination Airport, State Code

DestState CHAR(2) DEFAULT NULL,

-- 27 Destination Airport, State Fips

DestStateFips VARCHAR(10) DEFAULT NULL,

-- 28 Destination Airport, State Name

DestStateName VARCHAR(100) DEFAULT NULL,

-- 29 Destination Airport, World Area Code

DestWac DECIMAL(8,2) DEFAULT NULL,

-- 30 CRS Departure Time (local time: hhmm)

CRSDepTime DECIMAL(8,2) DEFAULT NULL,

-- 31 Actual Departure Time (local time: hhmm)

DepTime DECIMAL(8,2) DEFAULT NULL,

-- 32 Difference in minutes between scheduled and actual departure time.

DepDelay DECIMAL(8,2) DEFAULT NULL,

-- 33 Difference in minutes between scheduled and actual departure time.

DepDelayMinutes DECIMAL(8,2) DEFAULT NULL,

-- 34 Departure Delay Indicator, 15 Minutes or More (1=Yes)

DepDel15 DECIMAL(8,2) DEFAULT NULL,

-- 35 Departure Delay intervals, every (15 minutes from <-15 to >180)

DepartureDelayGroups DECIMAL(8,2) DEFAULT NULL,

-- 36 CRS Departure Time Block, Hourly Intervals

DepTimeBlk VARCHAR(20) DEFAULT NULL,

-- 37 Taxi Out Time, in Minutes

TaxiOut DECIMAL(8,2) DEFAULT NULL,

-- 38 Wheels Off Time (local time: hhmm)

WheelsOff DECIMAL(8,2) DEFAULT NULL,

-- 39 Wheels On Time (local time: hhmm)

WheelsOn DECIMAL(8,2) DEFAULT NULL,

-- 40 Taxi In Time, in Minutes

TaxiIn DECIMAL(8,2) DEFAULT NULL,

-- 41 CRS Arrival Time (local time: hhmm)

CRSArrTime DECIMAL(8,2) DEFAULT NULL,

-- 42 Actual Arrival Time (local time: hhmm)

ArrTime DECIMAL(8,2) DEFAULT NULL,

-- 43 Difference in minutes between scheduled and actual arrival time.

ArrDelay DECIMAL(8,2) DEFAULT NULL,

-- 44 Difference in minutes between scheduled and actual arrival time.

ArrDelayMinutes DECIMAL(8,2) DEFAULT NULL,

-- 45 Arrival Delay Indicator, 15 Minutes or More (1=Yes)

ArrDel15 DECIMAL(8,2) DEFAULT NULL,

-- 46 Arrival Delay intervals, every (15-minutes from <-15 to >180)

ArrivalDelayGroups DECIMAL(8,2) DEFAULT NULL,

-- 47 CRS Arrival Time Block, Hourly Intervals

ArrTimeBlk VARCHAR(20) DEFAULT NULL,

-- 48 Cancelled Flight Indicator (1=Yes)

Cancelled TINYINT DEFAULT NULL,

-- 49 Specifies The Reason For Cancellation

CancellationCode CHAR(1),

-- 50 Diverted Flight Indicator (1=Yes)

Diverted TINYINT DEFAULT NULL,

-- 51 CRS Elapsed Time of Flight, in Minutes

CRSElapsedTime DECIMAL(8,2) DEFAULT NULL,

-- 52 Elapsed Time of Flight, in Minutes

ActualElapsedTime DECIMAL(8,2) DEFAULT NULL,

-- 53 Flight Time, in Minutes

AirTime DECIMAL(8,2) DEFAULT NULL,

-- 54 Number of Flights

Flights DECIMAL(8,2) DEFAULT NULL,

-- 55 Distance between Airports (miles)

Distance DECIMAL(8,2) DEFAULT NULL,

-- 56 Distance Intervals, every 250 Miles, for Flight Segment

DistanceGroup TINYINT DEFAULT NULL,

-- 57 Carrier Delay, in Minutes

CarrierDelay DECIMAL(8,2) DEFAULT NULL,

-- 58 Weather Delay, in Minutes

WeatherDelay DECIMAL(8,2) DEFAULT NULL,

-- 59 National Air System Delay, in Minutes

NASDelay DECIMAL(8,2) DEFAULT NULL,

-- 60 Security Delay, in Minutes

SecurityDelay DECIMAL(8,2) DEFAULT NULL,

-- 61 Late Aircraft Delay, in Minutes

LateAircraftDelay DECIMAL(8,2) DEFAULT NULL,

-- 62 First Gate Departure Time at Origin Airport

FirstDepTime VARCHAR(10) DEFAULT NULL,

-- 63 Total Ground Time Away from Gate for Gate Return or Cancelled Flight

TotalAddGTime VARCHAR(10) DEFAULT NULL,

-- 64 Longest Time Away from Gate for Gate Return or Cancelled Flight

LongestAddGTime VARCHAR(10) DEFAULT NULL,

-- 65 Number of Diverted Airport Landings

DivAirportLandings VARCHAR(10) DEFAULT NULL,

-- 66 Diverted Flight Reaching Scheduled Destination Indicator (1=Yes)

DivReachedDest VARCHAR(10) DEFAULT NULL,

-- 67 Elapsed Time of Diverted Flight Reaching Scheduled Destination, in Minutes

DivActualElapsedTime VARCHAR(10) DEFAULT NULL,

-- 68 Difference in minutes between scheduled and actual arrival time for a diverted flight …

DivArrDelay VARCHAR(10) DEFAULT NULL,

-- 69 Distance between scheduled destination and final diverted airport (miles).

DivDistance VARCHAR(10) DEFAULT NULL,

-- 70 Diverted Airport Code1

Div1Airport VARCHAR(10) DEFAULT NULL,

-- 71 Airport ID of Diverted Airport 1.

Div1AirportID VARCHAR (10) DEFAULT NULL,

-- 72 Airport Sequence ID of Diverted Airport 1.

Div1AirportSeqID VARCHAR (10) DEFAULT NULL,

-- 73 Wheels On Time (local time: hhmm) at Diverted Airport Code1

Div1WheelsOn VARCHAR(10) DEFAULT NULL,

-- 74 Total Ground Time Away from Gate at Diverted Airport Code1

Div1TotalGTime VARCHAR (10) DEFAULT NULL,

-- 75 Longest Ground Time Away from Gate at Diverted Airport Code1

Div1LongestGTime VARCHAR (10) DEFAULT NULL,

-- 76 Wheels Off Time (local time: hhmm) at Diverted Airport Code1

Div1WheelsOff VARCHAR (10) DEFAULT NULL,

-- 77 Aircraft Tail Number for Diverted Airport Code1

Div1TailNum VARCHAR (10) DEFAULT NULL,

-- 78 Diverted Airport Code2

Div2Airport VARCHAR (10) DEFAULT NULL,

-- 79 Airport ID of Diverted Airport 2.

Div2AirportID VARCHAR(10) DEFAULT NULL,

-- 80 Airport Sequence ID of Diverted Airport 2.

Div2AirportSeqID VARCHAR (10) DEFAULT NULL,

-- 81 Wheels On Time (local time: hhmm) at Diverted Airport Code2

Div2WheelsOn VARCHAR (10) DEFAULT NULL,

-- 82 Total Ground Time Away from Gate at Diverted Airport Code2

Div2TotalGTime VARCHAR (10) DEFAULT NULL,

-- 83 Longest Ground Time Away from Gate at Diverted Airport Code2

Div2LongestGTime VARCHAR (10) DEFAULT NULL,

-- 84 Wheels Off Time (local time: hhmm) at Diverted Airport Code2

Div2WheelsOff VARCHAR (10) DEFAULT NULL,

-- 85 Aircraft Tail Number for Diverted Airport Code2

Div2TailNum VARCHAR (10) DEFAULT NULL,

-- 86 Diverted Airport Code3

Div3Airport VARCHAR (10) DEFAULT NULL,

-- 87 Airport ID of Diverted Airport 3.

Div3AirportID VARCHAR(10) DEFAULT NULL,

-- 88 Airport Sequence ID of Diverted Airport 3.

Div3AirportSeqID VARCHAR (10) DEFAULT NULL,

-- 89 Wheels On Time (local time: hhmm) at Diverted Airport Code3

Div3WheelsOn VARCHAR (10) DEFAULT NULL,

-- 90 Total Ground Time Away from Gate at Diverted Airport Code3

Div3TotalGTime VARCHAR (10) DEFAULT NULL,

-- 91 Longest Ground Time Away from Gate at Diverted Airport Code3

Div3LongestGTime VARCHAR (10) DEFAULT NULL,

-- 92 Wheels Off Time (local time: hhmm) at Diverted Airport Code3

Div3WheelsOff VARCHAR (10) DEFAULT NULL,

-- 93 Aircraft Tail Number for Diverted Airport Code3

Div3TailNum VARCHAR (10) DEFAULT NULL,

-- 94 Diverted Airport Code4

Div4Airport VARCHAR (10) DEFAULT NULL,

-- 95 Airport ID of Diverted Airport 4.

Div4AirportID VARCHAR(10) DEFAULT NULL,

-- 96 Airport Sequence ID of Diverted Airport 4.

Div4AirportSeqID VARCHAR (10) DEFAULT NULL,

-- 97 Wheels On Time (local time: hhmm) at Diverted Airport Code4

Div4WheelsOn VARCHAR (10) DEFAULT NULL,

-- 98 Total Ground Time Away from Gate at Diverted Airport Code4

Div4TotalGTime VARCHAR (10) DEFAULT NULL,

-- 99 Longest Ground Time Away from Gate at Diverted Airport Code4

Div4LongestGTime VARCHAR (10) DEFAULT NULL,

-- 100 Wheels Off Time (local time: hhmm) at Diverted Airport Code4

Div4WheelsOff VARCHAR (10) DEFAULT NULL,

-- 101 Aircraft Tail Number for Diverted Airport Code4

Div4TailNum VARCHAR (10) DEFAULT NULL,

-- 102 Diverted Airport Code5

Div5Airport VARCHAR (10) DEFAULT NULL,

-- 103 Airport ID of Diverted Airport 5.

Div5AirportID VARCHAR(10) DEFAULT NULL,

-- 104 Airport Sequence ID of Diverted Airport 5.

Div5AirportSeqID VARCHAR (10) DEFAULT NULL,

-- 105 Wheels On Time (local time: hhmm) at Diverted Airport Code5

Div5WheelsOn VARCHAR (10) DEFAULT NULL,

-- 106 Total Ground Time Away from Gate at Diverted Airport Code5

Div5TotalGTime VARCHAR (10) DEFAULT NULL,

-- 107 Longest Ground Time Away from Gate at Diverted Airport Code5

Div5LongestGTime VARCHAR (10) DEFAULT NULL,

-- 108 Wheels Off Time (local time: hhmm) at Diverted Airport Code5

Div5WheelsOff VARCHAR (10) DEFAULT NULL,

-- 109 Aircraft Tail Number for Diverted Airport Code5

Div5TailNum VARCHAR (10) DEFAULT NULL

);

1. http://www.transtats.bts.gov/ [↑](#footnote-ref-1)
2. http://www.transtats.bts.gov/DL\_SelectFields.asp?Table\_ID=236&DB\_Short\_Name=On-Time [↑](#footnote-ref-2)