

```

# $ sudo lsb_release -d
#   Description:      Ubuntu 15.10
#
# $ sudo cat /proc/cpuinfo | grep model\ name
#   model name       : Intel(R) Pentium(R) CPU 2020M @ 2.40GHz
#   model name       : Intel(R) Pentium(R) CPU 2020M @ 2.40GHz
#
# $ sudo free -h
#              total        used        free      shared    buffers     cached
#   Mem:        3.7G         3.4G         338M         278M         46M         2.5G
#   -/+ buffers/cache:      934M         2.8G
#   Swap:        1.9G         955M         952M
#
# $ sudo lshw -short -class disk | grep sda
#   /0/2/0.0.0      /dev/sda    disk              500GB Hitachi HTS54505
#
# $ mysql -u root -p --local-infile

# Q-1 - Import the datasets.

SHOW DATABASES;
USE assignments;
CREATE TABLE ontime (
    Year INT, Month INT, DayOfMonth INT, DayOfWeek INT, DepTime INT,
    CRSDepTime INT, ArrTime INT, CRSArrTime INT, UniqueCarrier VARCHAR(5),
    FlightNum INT, TailNum VARCHAR(8), ActualElapsedTime INT, CRSElapsedTime INT,
    AirTime INT, ArrDelay INT, DepDelay INT, Origin VARCHAR(3), Dest VARCHAR(3),
    Distance INT, TaxiIn INT, TaxiOut INT, Cancelled INT, CancellationCode VARCHAR(1),
    Diverted VARCHAR(1), CarrierDelay INT, WeatherDelay INT, NASDelay INT,
    SecurityDelay INT, LateAircraftDelay INT
);
SHOW FULL COLUMNS FROM ontime;
LOAD DATA LOCAL INFILE '/tmp/2004.csv'
    INTO TABLE ontime
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 7129270 rows affected, 65535 warnings (3 min 7.46 sec)
-- Records: 7129270 Deleted: 0 Skipped: 0 Warnings: 821678
SHOW WARNINGS LIMIT 10;
LOAD DATA LOCAL INFILE '/tmp/2005.csv'
    INTO TABLE ontime
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 7140596 rows affected, 65535 warnings (3 min 10.90 sec)
-- Records: 7140596 Deleted: 0 Skipped: 0 Warnings: 858492
SHOW WARNINGS LIMIT 10;
LOAD DATA LOCAL INFILE '/tmp/2006.csv'
    INTO TABLE ontime
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 7141922 rows affected, 65535 warnings (3 min 12.79 sec)
-- Records: 7141922 Deleted: 0 Skipped: 0 Warnings: 796352
SHOW WARNINGS LIMIT 10;

```

```

LOAD DATA LOCAL INFILE '/tmp/2007.csv'
    INTO TABLE ontime
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 7453215 rows affected, 65535 warnings (3 min 22.57 sec)
-- Records: 7453215 Deleted: 0 Skipped: 0 Warnings: 1034198
SHOW WARNINGS LIMIT 10;
LOAD DATA LOCAL INFILE '/tmp/2008.csv'
    INTO TABLE ontime
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 7009728 rows affected, 65535 warnings (3 min 25.98 sec)
-- Records: 7009728 Deleted: 0 Skipped: 0 Warnings: 28602754
SHOW WARNINGS LIMIT 10;
ALTER TABLE ontime ADD INDEX Year (Year);
-- Query OK, 0 rows affected (5 min 30.55 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
ALTER TABLE ontime ADD INDEX Date (Year, Month, DayofMonth);
-- Query OK, 0 rows affected (9 min 18.92 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
ALTER TABLE ontime ADD INDEX Origin (Origin);
-- Query OK, 0 rows affected (6 min 7.66 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
ALTER TABLE ontime ADD INDEX Dest (Dest);
-- Query OK, 0 rows affected (6 min 38.85 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
CREATE TABLE plane (
    TailNum VARCHAR(7), Type VARCHAR(19), Manufacturer VARCHAR(30), IssueDate DATE,
    Model VARCHAR(17), Status VARCHAR(17), AircraftType VARCHAR(24),
    EngineType VARCHAR(13), Year INT
);
SHOW FULL COLUMNS FROM plane;
LOAD DATA LOCAL INFILE '/tmp/plane-data.csv'
    INTO TABLE plane
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES
    (TailNum, Type, Manufacturer, @var, Model, Status, AircraftType, EngineType, Year)
    SET IssueDate = STR_TO_DATE(@var, '%m/%d/%Y');
-- Query OK, 5029 rows affected, 3989 warnings (1.66 sec)
-- Records: 5029 Deleted: 0 Skipped: 0 Warnings: 3989
SHOW WARNINGS LIMIT 10;
CREATE TABLE airport (
    Iata VARCHAR(4), Airport VARCHAR(41), City VARCHAR(33), State VARCHAR(2),
    Country VARCHAR(30), Lat NUMERIC(11, 8), Lon NUMERIC(12, 8)
);
SHOW FULL COLUMNS FROM airport;
LOAD DATA LOCAL INFILE '/tmp/airports.csv'
    INTO TABLE airport
    FIELDS TERMINATED BY ',' OPTIONALLY ENCLOSED BY '"'
    LINES TERMINATED BY '\n'
    IGNORE 1 LINES;
-- Query OK, 3376 rows affected (0.78 sec)
-- Records: 3376 Deleted: 0 Skipped: 0 Warnings: 0

```

Q0 - How many records are in the dataset?

```
SELECT COUNT(*) FROM ontime;
```

```
-- +-----+
-- | COUNT(*) |
-- +-----+
-- | 35874731 |
-- +-----+
-- 1 row in set (15.29 sec)
```

Q1 - Please list the number of all different routes (origin -> destination)
and contain "DISTINCT" syntax in the usage of SQL.

```
SELECT COUNT(*) FROM (SELECT DISTINCT Origin, Dest FROM ontime) AS temp;
```

```
-- +-----+
-- | COUNT(*) |
-- +-----+
-- |      6901 |
-- +-----+
-- 1 row in set (2 min 1.59 sec)
```

Q2 - Please list the number of flights whose destination is JFK
and the actual flight time is between 1 to 3 hours
and contain "BETWEEN" syntax in the usage of SQL.

```
SELECT COUNT(*) FROM ontime WHERE
    Dest = 'JFK' AND (ActualElapsedTime BETWEEN 1*60 AND 3*60);
```

```
-- +-----+
-- | COUNT(*) |
-- +-----+
-- | 303868 |
-- +-----+
-- 1 row in set (36.54 sec)
```

Q3 - Please list both the aircraft type names and the amounts
whose aircraft type name begins with "737"
and contain "LIKE" syntax in the usage of SQL.

```
SELECT Model, COUNT(Model) FROM plane WHERE Model LIKE '737%' GROUP BY Model;
```

```
-- +-----+-----+
-- | Model      | COUNT(Model) |
-- +-----+-----+
-- | 737-230    | 1 |
-- | 737-236    | 4 |
-- | 737-282    | 1 |
-- | 737-282C   | 1 |
-- | 737-2P6    | 1 |
-- | 737-2X6C   | 1 |
-- | 737-2Y5    | 1 |
-- | 737-301    | 7 |
-- | 737-317    | 4 |
-- | 737-322    | 64 |
-- | 737-33A    | 4 |
-- | 737-3A4    | 9 |
-- | 737-3B7    | 16 |
-- | 737-3G7    | 20 |
-- | 737-3H4    | 147 |
-- | 737-3K2    | 2 |
-- | 737-3L9    | 2 |
```

```

-- | 737-3Q8 | 9 |
-- | 737-3S3 | 4 |

-- | 737-3T5 | 3 |
-- | 737-3T0 | 50 |
-- | 737-3Y0 | 6 |
-- | 737-401 | 13 |
-- | 737-490 | 16 |
-- | 737-4B7 | 27 |
-- | 737-4Q8 | 22 |
-- | 737-4S3 | 1 |
-- | 737-522 | 30 |
-- | 737-524 | 56 |
-- | 737-5H4 | 25 |
-- | 737-705 | 1 |
-- | 737-724 | 36 |
-- | 737-73A | 2 |
-- | 737-76N | 27 |
-- | 737-76Q | 2 |
-- | 737-790 | 19 |
-- | 737-7AD | 1 |
-- | 737-7BD | 32 |
-- | 737-7H4 | 308 |
-- | 737-7Q8 | 2 |
-- | 737-824 | 100 |
-- | 737-832 | 68 |
-- | 737-890 | 33 |
-- | 737-8FH | 1 |
-- | 737-924 | 12 |
-- | 737-924ER | 8 |
-- | 737-990 | 12 |
-- +-----+-----+
-- 47 rows in set (0.00 sec)

```

```

# Q4 - Please list the average speed (using the actual arrival time)
#       of all aircraft types and contain "AVG" in the usage of SQL.
#
# The data are too huge for my computer to process.
# I had executed this query for more than 30 minutes
# and the system resources were almost used up.

```

```

ALTER TABLE ontime ADD INDEX TailNum (TailNum);
-- Query OK, 0 rows affected (7 min 27.65 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
ALTER TABLE plane ADD INDEX TailNum (TailNum);
-- Query OK, 0 rows affected (0.76 sec)
-- Records: 0 Duplicates: 0 Warnings: 0
SELECT plane.Model, AVG(ontime.Distance/ontime.ActualElapsedTime*60) AS AvgSpeed
      FROM plane INNER JOIN ontime ON plane.TailNum = ontime.TailNum
      GROUP BY plane.Model;
-- ^Cctrl-C -- sending "KILL QUERY 5" to server ...
-- Ctrl-C -- query aborted.
-- ERROR 1317 (70100): Query execution was interrupted

```

```

# Q5 - Please list the top 50 timezone counts and routes (origin -> destination)
#       by the timezone count of routes (assume one timezone is formed by 15 degree
#       longitude and is not related to countries) in decreasing order
#       and contain "ORDER BY" in the usage of SQL.

```

```

#
# The data are too huge for my computer to process.
# I had executed this query for more than 30 minutes
# and the system resources were almost used up.

SELECT DISTINCT
    FLOOR(ABS(temp1.Lon-temp2.Lon)/15) AS Timezone, ontime.Origin, ontime.Dest
FROM ontime
INNER JOIN airport AS temp1 ON ontime.Origin = temp1.Iata
INNER JOIN airport AS temp2 ON ontime.Dest = temp2.Iata
ORDER BY Timezone DESC, ontime.Origin ASC, ontime.Dest ASC
LIMIT 0, 50;
-- ^Cctrl-C -- sending "KILL QUERY 5" to server ...
-- Ctrl-C -- query aborted.
-- ERROR 1317 (70100): Query execution was interrupted

# Q6 - Please list the aircraft types which never appeared after 2008 (appeared before)
#       and contain "NOT IN" syntax in the usage of SQL.
#
# Please note that not all plane data are in the "plane" table.
# So it is possible to receive nothing
# when we tries to find out their aircraft type names.

SELECT DISTINCT plane.Model FROM plane
    INNER JOIN ontime ON plane.TailNum = ontime.TailNum
WHERE plane.Model NOT IN (
    SELECT DISTINCT plane.Model FROM plane
    INNER JOIN ontime ON plane.TailNum = ontime.TailNum
    WHERE ontime.Year >= 2008
);
-- ^Cctrl-C -- sending "KILL QUERY 5" to server ...
-- Ctrl-C -- query aborted.
-- ERROR 1317 (70100): Query execution was interrupted
SELECT DISTINCT Model FROM plane WHERE TailNum IN (
    SELECT DISTINCT TailNum FROM ontime WHERE TailNum NOT IN (
        SELECT DISTINCT TailNum FROM ontime WHERE Year >= 2008
    )
);
-- Empty set (1 min 55.69 sec)

# Q7 - Please list the manufacturers whose airplanes' delay times
#       are more than half an hour.
#
# The data are too huge for my computer to process.
# I had executed this query for more than 30 minutes
# and the system resources were almost used up.

SELECT plane.Manufacturer, AVG(ontime.ArrDelay)
FROM plane INNER JOIN ontime ON plane.TailNum = ontime.TailNum
GROUP BY plane.Manufacturer
HAVING AVG(ontime.ArrDelay) > 30;
-- ^Cctrl-C -- sending "KILL QUERY 5" to server ...
-- Ctrl-C -- query aborted.
-- ERROR 1317 (70100): Query execution was interrupted

# Q8 - Please list the average delay time of a week (7 days) and
#       aday (24 hours), and when is the best time to fly
#       (having the minimum delay).

```

```
#
# The best day is Saturday (day 6) and the best hour is 5 o'clock.
```

```
SELECT DayOfWeek, AVG(ArrDelay) FROM ontime GROUP BY DayOfWeek;
```

```
-- +-----+-----+
-- | DayOfWeek | AVG(ArrDelay) |
-- +-----+-----+
-- |          1 |          8.4854 |
-- |          2 |          6.4295 |
-- |          3 |          7.6126 |
-- |          4 |          9.8577 |
-- |          5 |         10.5425 |
-- |          6 |          4.4603 |
-- |          7 |          7.9931 |
-- +-----+-----+
-- 7 rows in set (1 min 52.20 sec)
```

```
SELECT FLOOR(CRSDepTime/100) AS Hour, AVG(ArrDelay) FROM ontime GROUP BY Hour;
```

```
-- +-----+-----+
-- | Hour | AVG(ArrDelay) |
-- +-----+-----+
-- |    0 |          2.6160 |
-- |    1 |          1.6824 |
-- |    2 |          0.9439 |
-- |    3 |          8.8009 |
-- |    4 |          4.0973 |
-- |    5 |         -0.1883 |
-- |    6 |          0.1889 |
-- |    7 |          1.2136 |
-- |    8 |          2.4943 |
-- |    9 |          3.4210 |
-- |   10 |          4.4658 |
-- |   11 |          5.3243 |
-- |   12 |          6.4725 |
-- |   13 |          8.4028 |
-- |   14 |          9.9337 |
-- |   15 |         11.1849 |
-- |   16 |         12.3410 |
-- |   17 |         14.1203 |
-- |   18 |         14.4066 |
-- |   19 |         14.6210 |
-- |   20 |         14.3945 |
-- |   21 |         12.7192 |
-- |   22 |          8.7032 |
-- |   23 |          6.3922 |
-- +-----+-----+
-- 24 rows in set (1 min 49.76 sec)
```

```
# Q9 - Do older planes suffer more delay?
```

```
#
```

```
# No. As the following tables show, there is no positive correlation
# between the age of the plane and its delay.
```

```
SELECT DATEDIFF(IssueDate, CURDATE()) FROM plane WHERE TailNum IN (
    SELECT TailNum FROM ontime ORDER BY ArrDelay DESC LIMIT 10
);
```

```
-- ERROR 1235 (42000): This version of MySQL doesn't yet support
-- 'LIMIT & IN/ALL/ANY/SOME subquery'
```

```

SELECT DATEDIFF(IssueDate, CURDATE()) FROM plane WHERE TailNum IN (
    SELECT * FROM (
        SELECT TailNum FROM ontime ORDER BY ArrDelay DESC LIMIT 10
    ) AS temp
);

```

```

-- +-----+
-- | DATEDIFF(IssueDate, CURDATE()) |
-- +-----+
-- | -8741 |
-- | -7800 |
-- | -8675 |
-- | -6304 |
-- | -3293 |
-- | -3231 |
-- +-----+
-- 6 rows in set (1 min 45.33 sec)

```

```

SELECT DATEDIFF(IssueDate, CURDATE()) FROM plane WHERE TailNum IN (
    SELECT * FROM (
        SELECT TailNum FROM ontime ORDER BY ArrDelay ASC LIMIT 10
    ) AS temp
);

```

```

-- +-----+
-- | DATEDIFF(IssueDate, CURDATE()) |
-- +-----+
-- | -8200 |
-- | -7598 |
-- | -7352 |
-- | -7218 |
-- | -4373 |
-- | -4349 |
-- | -4608 |
-- | -4533 |
-- +-----+
-- 8 rows in set (1 min 11.22 sec)

```

Q10 - How well does weather or season predict plane delay?

#

Let's find the correlation between arrival delay and weather delay.

The correlation coefficient is 0.29, which means a weak positive correlation.

```

SELECT
    @AvgArrDelay := AVG(ArrDelay), @AvgWeatherDelay := AVG(WeatherDelay),
    @StdArrDelay := STDDEV_SAMP(ArrDelay), @StdWeatherDelay := STDDEV_SAMP(WeatherDelay)
FROM ontime;

```

```

-- +-----+-----+-----+-----+
-- | @AvgArrDelay | @AvgWeatherDelay | @StdArrDelay | @StdWeatherDelay |
-- +-----+-----+-----+-----+
-- | 7.990401238 | 0.696268886 | 36.12799877414426 | 8.868056284288038 |
-- +-----+-----+-----+-----+
-- 1 row in set (57.54 sec)

```

```

SELECT
    SUM( (ArrDelay-@AvgArrDelay)*(WeatherDelay-@AvgWeatherDelay) ) /
    ( (COUNT(ArrDelay)-1)*@StdArrDelay*@StdWeatherDelay )
FROM ontime;

```

```

-- +-----+
-- | SUM( (ArrDelay-@AvgArrDelay)*(WeatherDelay-... |
-- +-----+

```

```

-- | 0.2854138548147738 |
-- +-----+
-- 1 row in set (1 min 5.08 sec)

# Q11 - Can you detect cascaded failures
#       as delay in one airport causes delay in another?
#       Are there any critical links in the system?
#
# This plane went through DEN -> SEA -> SFO -> RNO on 2004-09-01.
# However, the first flight was seriously delayed,
# leading to cascaded delay in the second and the third flight.
# (Although the duration of the second and the third flight is as scheduled.)

SELECT CRSDepTime, DepTime, DepDelay, CRSArrTime, ArrTime, ArrDelay
      FROM assignments.ontime
      WHERE Year = 2004 AND Month = 9 AND DayOfMonth = 1
            AND TailNum = 'N530UA' AND LateAircraftDelay > 0;

-- +-----+-----+-----+-----+-----+-----+
-- | CRSDepTime | DepTime | DepDelay | CRSArrTime | ArrTime | ArrDelay |
-- +-----+-----+-----+-----+-----+-----+
-- |      1434 |    1623 |      109 |      1612 |    1803 |      111 |
-- |      1712 |    1846 |       94 |      1908 |    2042 |       94 |
-- |      2008 |    2114 |       66 |      2058 |    2204 |       66 |
-- +-----+-----+-----+-----+-----+-----+
-- 3 rows in set (0.02 sec)

SELECT CRSElapsedTime, ActualElapsedTime, LateAircraftDelay, Origin, Dest
      FROM assignments.ontime
      WHERE Year = 2004 AND Month = 9 AND DayOfMonth = 1
            AND TailNum = 'N530UA' AND LateAircraftDelay > 0;

-- +-----+-----+-----+-----+-----+-----+
-- | CRSElapsedTime | ActualElapsedTime | LateAircraftDelay | Origin | Dest |
-- +-----+-----+-----+-----+-----+-----+
-- |           158 |           160 |           106 | DEN    | SEA  |
-- |           116 |           116 |            94 | SEA    | SFO  |
-- |            50 |            50 |            66 | SFO    | RNO  |
-- +-----+-----+-----+-----+-----+-----+
-- 3 rows in set (0.02 sec)

# Q12 - Feel free to think.
#       Write down any valuable observation with explanation.
#
# The most convenient airport in the United States in 2008
# is William B Hartsfield-Atlanta Intl in Atlanta, GA.
# We can head for 173 different airports there.

CREATE TEMPORARY TABLE IF NOT EXISTS temp AS (
      SELECT Origin, COUNT(DISTINCT Dest) AS DestCount
      FROM ontime WHERE Year = 2008 GROUP BY Origin
);
-- Query OK, 303 rows affected (1 min 33.62 sec)
-- Records: 303 Duplicates: 0 Warnings: 0

SELECT airport.Airport AS AirportName,
      CONCAT(airport.City, ', ', airport.State) AS Location,
      temp.DestCount AS AccessibleAirportCount
      FROM airport
      INNER JOIN temp ON airport.Iata = temp.Origin
      ORDER BY AccessibleAirportCount DESC
      LIMIT 0, 50;

```


+-----+-----+-----+			
-- AirportName		Location	Accessible...
+-----+-----+-----+			
-- William B Hartsfield-Atlanta Intl	Atlanta, GA		173
-- Chicago O'Hare International	Chicago, IL		149
-- Dallas-Fort Worth International	Dallas-Fort Worth, TX		134
-- Denver Intl	Denver, CO		127
-- Minneapolis-St Paul Intl	Minneapolis, MN		126
-- Detroit Metropolitan-Wayne County	Detroit, MI		118
-- George Bush Intercontinental	Houston, TX		114
-- Salt Lake City Intl	Salt Lake City, UT		114
-- Cincinnati Northern Kentucky Intl	Covington, KY		113
-- Newark Intl	Newark, NJ		92
-- McCarran International	Las Vegas, NV		91
-- Los Angeles International	Los Angeles, CA		90
-- Orlando International	Orlando, FL		89
-- Phoenix Sky Harbor International	Phoenix, AZ		88
-- Charlotte/Douglas International	Charlotte, NC		82
-- Memphis International	Memphis, TN		79
-- Cleveland-Hopkins Intl	Cleveland, OH		75
-- San Francisco International	San Francisco, CA		74
-- Washington Dulles International	Chantilly, VA		71
-- John F Kennedy Intl	New York, NY		68
-- Baltimore-Washington International	Baltimore, MD		65
-- Gen Edw L Logan Intl	Boston, MA		63
-- Tampa International	Tampa, FL		62
-- Philadelphia Intl	Philadelphia, PA		61
-- LaGuardia	New York, NY		60
-- Fort Lauderdale-Hollywood Int'l	Ft. Lauderdale, FL		58
-- Seattle-Tacoma Intl	Seattle, WA		56
-- Chicago Midway	Chicago, IL		54
-- San Diego International-Lindbergh	San Diego, CA		54
-- General Mitchell International	Milwaukee, WI		53
-- Ronald Reagan Washington National	Arlington, VA		52
-- Kansas City International	Kansas City, MO		52
-- Austin-Bergstrom International	Austin, TX		52
-- Lambert-St Louis International	St Louis, MO		50
-- Nashville International	Nashville, TN		49
-- Miami International	Miami, FL		48
-- Portland Intl	Portland, OR		45
-- San Antonio International	San Antonio, TX		43
-- Raleigh-Durham International	Raleigh, NC		43
-- New Orleans International	New Orleans, LA		41
-- Indianapolis International	Indianapolis, IN		40
-- Albuquerque International	Albuquerque, NM		38
-- Port Columbus Intl	Columbus, OH		38
-- Sacramento International	Sacramento, CA		37
-- Southwest Florida International	Ft. Myers, FL		35
-- Jacksonville International	Jacksonville, FL		35
-- Metropolitan Oakland International	Oakland, CA		34
-- Pittsburgh International	Pittsburgh, PA		34
-- William P Hobby	Houston, TX		33
-- Bradley International	Windsor Locks, CT		33
+-----+-----+-----+			
-- 50 rows in set (0.11 sec)			

幫助教訂正英文有加分嗎 ^__^

這份題目裡的英文幾乎沒有一句是寫對的欸QQ

幾乎所有關係代名詞跟名詞單複數都用錯...