"Oh, oh big ol' jet airliner - don't carry me too far away"

The airline industry has in recent years taken a huge nosedive in popularity. Rising fuel costs and a razor thin profit margin has led the industry to implement several new fees (bags, blankets, pillows, food, drinks) that have generated billions of dollars in profit. However, flights are still being cut, airports are being closed, and airlines are constantly merging in an effort to keep their companies profitable. Both anecdotal evidence and surveys continue to show though that passengers are still likely to be angry about arriving late to their destination. Delays can lead to missed flights and inconvenience and expense for both the traveler who is displaced and the airline who often must cover the costs. A better understanding of what contributes to arrival delays is crucial.

The Bureau of Transportation Statistics has been keeping track of every flight over at least the past twenty years and has collected data on every segment of the flight. The entire flight can be broken down into: leaving the gate and taxiing to the runway, the actual flight time, and taxiing to the gate upon landing. All three of these segments contribute to whether or not a flight is delayed. The BTS knows the flight's scheduled departure and arrival time (denoted with a CRS) but also the actual departure and arrival time. Delays can occur due to weather, the National Air System (NAS), security problems, late aircraft, or the carrier/company itself. Flights are also labeled by their unique carrier (e.g., American Airlines), their flight number, and their tail number (corresponds to plane type). See `http://www.transtats.bts.gov/Fields.asp?Table_ID=236` for explanations of the types of variables the Bureau tracks. More details about the airports and the carriers can be found in `airports.csv` and `carriers.csv`.

In 2008, there were 7,009,728 flights tracked by the BTS; your research group has been given a sample of these flights. (Note that you do not have flights that are continuation segments of longer flights. Each flight is likely a segment of a longer flight but your flights have been de-coupled.) Your research group has been asked by the Bureau of Transportation Statistics to analyze the data and develop a model to predict the arrival delay with the following variables:

ArrDelay: arrival delay in minutes (negative means an early arrival)

Month: month of the flight (1 = January; 12 = December)

DayofWeek: day of the week (1 = Monday; 7 = Sunday)

CRSDepTime: scheduled departure time (local time, hhmm)

CRSArrTime: scheduled arrival time (local time, hhmm)

AirTime: total time in the air in minutes

Distance: total distance of the flight in miles

TaxiIn: time taxiing out to the runway in minutes

TaxiOut: time taxiing to the gate after landing in minutes

DepDelay: departure delay in minutes (negative means an early departure)

Diverted: was the plane diverted? 1= Yes, 0 = No

Carrier: was there a delay due to the carrier/company? 1 = Yes, 0 = No

Weather: was there a delay due to weather? 1 = Yes, 0 = No

NAS: was there a delay due to the National Air System? 1 = Yes, 0 = No

Security: was there a delay due to security problems? 1 = Yes, 0 = No

LateAircraft: was there a delay due to a late aircraft? 1 = Yes, 0 = No