Use the site below for the data that will be needed for this part of this project:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/HG7NV7

Variable descriptions

	Name	**Description**
1	Year	1987-2008
2	Month	1-12
3	DayofMonth	1-31
4	DayOfWeek	1 (Monday) - 7 (Sunday)
5	DepTime	actual departure time (local, hhmm)
6	CRSDepTime	scheduled departure time (local, hhmm)
7	ArrTime	actual arrival time (local, hhmm)
8	CRSArrTime	scheduled arrival time (local, hhmm)
9	UniqueCarrier	unique carrier code
10	FlightNum	flight number
11	TailNum	plane tail number
12	ActualElapsedTime	in minutes
13	CRSElapsedTime	in minutes
14	AirTime	in minutes
15	ArrDelay	arrival delay, in minutes
16	DepDelay	departure delay, in minutes
17	Origin	origin IATA airport code
18	Dest	destination IATA airport code
19	Distance	in miles
20	TaxiIn	taxi in time, in minutes
21	TaxiOut	taxi out time in minutes
22	Cancelled	was the flight cancelled?
23	CancellationCode	reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
24	Diverted	1 = yes, 0 = no
25	CarrierDelay	in minutes
26	WeatherDelay	in minutes
27	NASDelay	in minutes
28	SecurityDelay	in minutes
29	LateAircraftDelay	in minutes

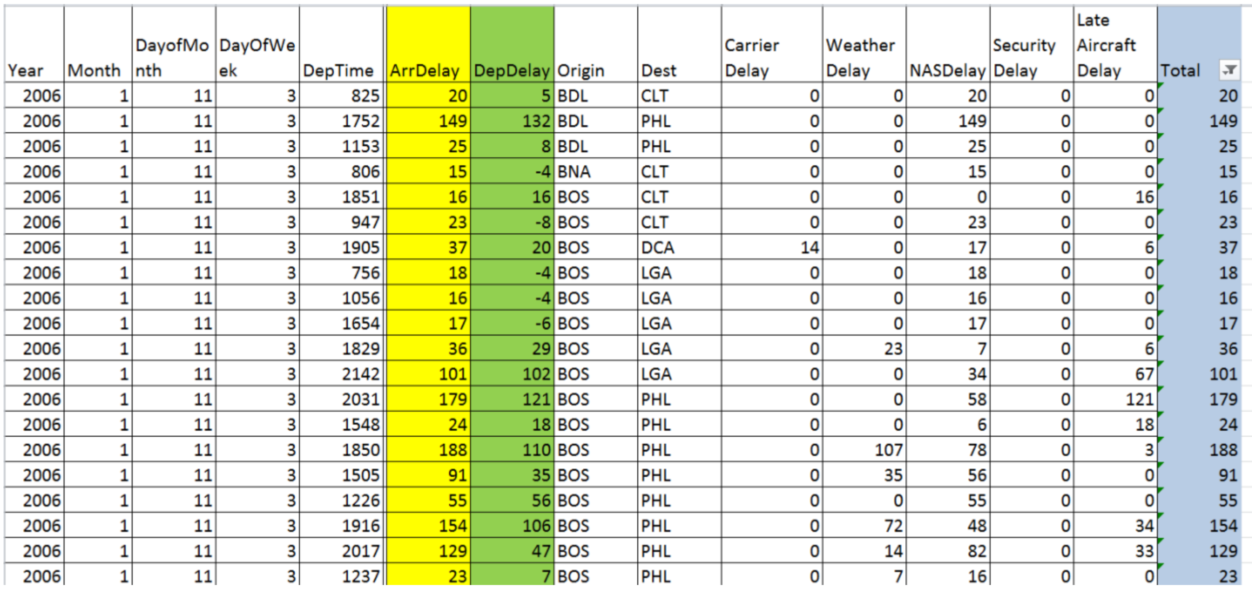
Note below site where you have auxilliary files on the site to convert the **carrier** code and **airport** names.

http://stat-computing.org/dataexpo/2009/supplemental-data.html

Each Person will be assigned a year and will do these tasks with the file

1. extract the data from the site and **download the cvs file from the website**
2. create a Hadoop table and load all the file data into it. Prefix the table name with your name.
3. display a sample of the loaded rows
4. determine the three **airports** with the highest delay time (in hours)
5. determine the three **carriers** with the highest delay time (in hours)
6. determine overall which type of delay (arrivals or departures) is the largest for **airports**

Example of some records in spreadsheet format:



Year	Month	DayofMonth	DayOfWeek	DepTime	ArrDelay	DepDelay	Origin	Dest	Carrier Delay	Weather Delay	NASDelay	Security Delay	Late Aircraft Delay	Total
2006	1	11	3	825	20	5	BDL	CLT	0	0	20	0	0	20
2006	1	11	3	1752	149	132	BDL	PHL	0	0	149	0	0	149
2006	1	11	3	1153	25	8	BDL	PHL	0	0	25	0	0	25
2006	1	11	3	806	15	-4	BNA	CLT	0	0	15	0	0	15
2006	1	11	3	1851	16	16	BOS	CLT	0	0	0	0	16	16
2006	1	11	3	947	23	-8	BOS	CLT	0	0	23	0	0	23
2006	1	11	3	1905	37	20	BOS	DCA	14	0	17	0	6	37
2006	1	11	3	756	18	-4	BOS	LGA	0	0	18	0	0	18
2006	1	11	3	1056	16	-4	BOS	LGA	0	0	16	0	0	16
2006	1	11	3	1654	17	-6	BOS	LGA	0	0	17	0	0	17
2006	1	11	3	1829	36	29	BOS	LGA	0	23	7	0	6	36
2006	1	11	3	2142	101	102	BOS	LGA	0	0	34	0	67	101
2006	1	11	3	2031	179	121	BOS	PHL	0	0	58	0	121	179
2006	1	11	3	1548	24	18	BOS	PHL	0	0	6	0	18	24
2006	1	11	3	1850	188	110	BOS	PHL	0	107	78	0	3	188
2006	1	11	3	1505	91	35	BOS	PHL	0	35	56	0	0	91
2006	1	11	3	1226	55	56	BOS	PHL	0	0	55	0	0	55
2006	1	11	3	1916	154	106	BOS	PHL	0	72	48	0	34	154
2006	1	11	3	2017	129	47	BOS	PHL	0	14	82	0	33	129
2006	1	11	3	1237	23	7	BOS	PHL	0	7	16	0	0	23