

DS 600 DATA MINING PROJECT PROPOSAL

BY

BHAGYASHREE

PROBLEM STATEMENT

Flight delayed can be huge problem to airline companies and in people's daily life. Therefore the study focus on which is the best airline companies' carrier are most reliable carrier. We will use Pattern analysis concept from Data mining

Sample data set

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	year	month	carrier	carrier_name	airport	airport_name	arr_flight	arr_del15	carrier_ct	weather	nas_ct	security	late_aircr	arr_canc	arr_divert
2	2004	1	DL	Delta Air Lines Inc.	PBI	West Palm Beach/Palm Beach, FL: Palm Beach International	650	126	21.06	6.44	51.58	1	45.92	4	0
3	2004	1	DL	Delta Air Lines Inc.	PDX	Portland, OR: Portland International	314	61	14.09	2.61	34.25	0	10.05	30	3
4	2004	1	DL	Delta Air Lines Inc.	PHL	Philadelphia, PA: Philadelphia International	513	97	27.6	0.42	51.86	0	17.12	15	0
5	2004	1	DL	Delta Air Lines Inc.	PHX	Phoenix, AZ: Phoenix Sky Harbor International	334	78	20.14	2.02	39.39	0	16.45	3	1
6	2004	1	DL	Delta Air Lines Inc.	PIT	Pittsburgh, PA: Pittsburgh International	217	47	8.08	0.44	21.89	0	16.59	4	1
7	2004	1	DL	Delta Air Lines Inc.	PNS	Pensacola, FL: Pensacola International	181	42	10.48	1.06	11.87	0	18.58	2	0
8	2004	1	DL	Delta Air Lines Inc.	PSC	Pasco/Kennewick/Richland, WA: Tri Cities	10	3	1.4	1	0.6	0	0	1	0
9	2004	1	DL	Delta Air Lines Inc.	PSP	Palm Springs, CA: Palm Springs International	31	2	1	0	0.46	0	0.54	0	0
10	2004	1	DL	Delta Air Lines Inc.	PVD	Providence, RI: Theodore Francis Green State	216	42	17	0.43	17.06	0	7.51	3	0
11	2004	1	DL	Delta Air Lines Inc.	PWM	Portland, ME: Portland International Jetport	122	21	8.14	0	8.06	0	4.8	2	0
12	2004	1	DL	Delta Air Lines Inc.	RDU	Raleigh/Durham, NC: Raleigh-Durham International	429	88	29.65	2.55	34.99	0	20.81	11	1
13	2004	1	DL	Delta Air Lines Inc.	RIC	Richmond, VA: Richmond International	239	48	16.51	0.07	13.62	0	17.79	12	0
14	2004	1	DL	Delta Air Lines Inc.	RSW	Fort Myers, FL: Southwest Florida International	437	97	23.57	4.59	44.55	0.69	23.59	1	0
15	2004	1	DL	Delta Air Lines Inc.	SAN	San Diego, CA: San Diego International	432	88	27.03	2.48	45.37	0	13.12	3	0
16	2004	1	DL	Delta Air Lines Inc.	SAT	San Antonio, TX: San Antonio International	277	72	12.72	1.29	34.07	0	23.92	6	0
17	2004	1	DL	Delta Air Lines Inc.	SAV	Savannah, GA: Savannah/Hilton Head International	263	46	9.93	1.03	22.26	0	12.77	4	0
18	2004	1	DL	Delta Air Lines Inc.	SDF	Louisville, KY: Louisville Muhammad Ali International	208	34	7.35	0.74	12.05	0	13.87	9	0
19	2004	1	DL	Delta Air Lines Inc.	SEA	Seattle, WA: Seattle/Tacoma International	470	109	17.74	1.88	73.78	0	15.6	7	1
20	2004	1	DL	Delta Air Lines Inc.	SFO	San Francisco, CA: San Francisco International	603	180	27.95	5.55	123.05	0	23.44	15	0
21	2004	1	DL	Delta Air Lines Inc.	SJC	San Jose, CA: Norman Y. Mineta San Jose International	145	36	9.12	0.9	17.94	0	8.05	1	0
22	2004	1	DL	Delta Air Lines Inc.	SJU	San Juan, PR: Luis Munoz Marin International	163	34	8.15	1.38	22.28	0	2.19	1	1
23	2004	1	DL	Delta Air Lines Inc.	SLC	Salt Lake City, UT: Salt Lake City International	3012	756	163.51	33.04	427.21	1.6	130.64	38	1
24	2004	1	DL	Delta Air Lines Inc.	SMF	Sacramento, CA: Sacramento International	155	48	12.8	0.55	26.17	0	8.48	0	0
25	2004	1	DL	Delta Air Lines Inc.	SNA	Santa Ana, CA: John Wayne Airport-Orange County	265	56	10.92	0.43	33.08	0	11.57	1	1
26	2004	1	DL	Delta Air Lines Inc.	SRQ	Sarasota/Bradenton, FL: Sarasota/Bradenton International	151	32	7.29	0	14.19	0	10.53	2	0
27	2004	1	DL	Delta Air Lines Inc.	STL	St. Louis, MO: St Louis Lambert International	228	52	18.06	0.91	22.71	0	10.31	18	0
28	2004	1	DL	Delta Air Lines Inc.	STT	Charlotte Amalie, VI: Cyril E King	56	11	0.35	0	9.19	0	1.46	0	0
29	2004	1	DL	Delta Air Lines Inc.	SYR	Syracuse, NY: Syracuse Hancock International	62	21	1.31	0.26	13.42	0	6.01	0	0
30	2004	1	DL	Delta Air Lines Inc.	TLH	Tallahassee, FL: Tallahassee International	154	23	8.37	1.06	8.72	0	4.85	1	0
31	2004	1	DL	Delta Air Lines Inc.	TPA	Tampa, FL: Tampa International	955	156	40.05	8.12	59.31	0	48.52	6	1
32	2004	1	DL	Delta Air Lines Inc.	TUL	Tulsa, OK: Tulsa International	104	30	7.59	1.05	16.36	0	5.01	2	1
33	2004	1	DL	Delta Air Lines Inc.	TUS	Tucson, AZ: Tucson International	62	14	3.04	0	7.1	0	3.86	1	0
34	2004	1	DL	Delta Air Lines Inc.	TVC	Knoxville, TN: McGhee Tyson	110	27	7.60	0.27	10.04	0	9.1	2	0

Attributes about data

1. Source is United States Department of Transportation
2. This is data set which has more than ~200,000 rows and 20 features.
3. The data has contained date, airport names, arrival flights count, delayed flights counts, late aircraft counts and so on.
4. Data set contain categorical information, and numeric information.

Goal and Task

More investigating insight fact about data

1. EDA will be performed on this data set.
2. Recommendation to carrier companies
3. Data mining tools will be applied on carrier company data set (classification)
4. Business values will be achieved from analysis.



Thank you
