

Name: Chandana Ramesh Galgali

Batch: FDS-2

Roll No: 16010422234

FDS Expt 1

1. Download a large dataset for the purpose of exploration and ensure that the dataset has a variety of attributes; number of attributes must be at least 25.

Source of data: Kaggle

Dataset: Flight Delay and Causes

No. of attributes: 28

Screenshots:

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	DayOfWeek	DepTime	ArrTime	CRSArrTime	UniqueCarrier	Airline	FlightNum	TailNum	ActualElapsedTime	CRSElapseTime	ArrDelay	DepDelay	Origin	Org_AirPort	Dest	AirPort	Distance	TaxiIn	TaxiOut	Cancelled	CancelledReason	Diverted	CarrierDelay	WeatherDelay	NASDelay	SecurityDelay	LateAircraft		
2	4	3/1/2019	1829	1959	1925	WN	Southwest	3920 N464WN	90	90	77	34	34	IND	Indianapolis	Baltimore	515	3	10	0	N	0	2	0	0	0	0	32	
3	4	3/1/2019	1937	2037	1940	WN	Southwest	509 N763SW	240	250	230	57	67	IND	Indianapolis	McCarran	1591	3	7	0	N	0	10	0	0	0	0	47	
4	4	3/1/2019	1644	1845	1725	WN	Southwest	1333 N334SW	121	135	107	80	94	IND	Indianapolis	Orlando Int	828	6	8	0	N	0	8	0	0	0	0	72	
5	4	3/1/2019	1452	1640	1625	WN	Southwest	675 N286WN	228	240	213	15	27	IND	Indianapolis	Phoenix SI	1489	7	8	0	N	0	3	0	0	0	0	12	
6	4	3/1/2019	1323	1526	1510	WN	Southwest	4 N674AA	123	135	110	16	28	IND	Indianapolis	TPA	Tampa Int	838	4	9	0	N	0	0	0	0	0	16	
7	4	3/1/2019	1416	1512	1435	WN	Southwest	54 N643SW	56	70	49	37	51	ISP	Long Island	BWI	Baltimore	220	2	5	0	N	0	12	0	0	0	25	
8	4	3/1/2019	1657	1754	1735	WN	Southwest	623 N724SW	57	70	47	19	32	ISP	Long Island	BWI	Baltimore	220	5	5	0	N	0	7	0	0	0	12	
9	4	3/1/2019	1422	1657	1610	WN	Southwest	188 N215WN	155	195	143	47	87	ISP	Long Island	FLL	Fort Lauderdale	1093	6	6	0	N	0	40	0	0	0	7	
10	4	3/1/2019	2107	2334	2230	WN	Southwest	362 N798SW	147	165	134	64	82	ISP	Long Island	MCO	Orlando Int	972	6	7	0	N	0	5	0	0	0	59	
11	4	3/1/2019	1812	1927	1815	WN	Southwest	422 N779SW	135	145	118	72	82	ISP	Long Island	MDW	Chicago Mid	765	6	11	0	N	0	3	0	0	0	69	
12	4	3/1/2019	1326	1559	1530	WN	Southwest	1056 N459WN	153	180	143	29	56	ISP	Long Island	PBI	Palm Beach	1052	5	5	0	N	0	0	0	0	0	29	
13	4	3/1/2019	1450	1806	1745	WN	Southwest	3244 N475WN	136	130	121	21	15	JAN	Jacksonville	BWI	Baltimore	888	/	8	0	N	0	0	0	0	0	13	
14	4	3/1/2019	2245	2354	1850	WN	Southwest	186 N792SW	69	80	59	304	315	JAN	Jacksonville	HOU	William P	359	3	7	0	N	0	282	0	0	0	22	
15	4	3/1/2019	2025	2135	2100	WN	Southwest	3154 N252WN	70	80	60	35	45	JAN	Jacksonville	HOU	William P	359	3	7	0	N	0	26	0	0	0	9	
16	4	3/1/2019	1038	1314	1225	WN	Southwest	1035 N346SW	96	100	81	49	53	JAN	Jacksonville	MCO	Orlando Int	587	8	7	0	N	0	7	0	0	0	42	
17	4	3/1/2019	1900	2123	2045	WN	Southwest	205 N299WN	143	115	97	38	10	JAN	Jacksonville	MDW	Chicago Mid	666	40	6	0	N	0	1	0	28	0	9	
18	4	3/1/2019	948	959	940	WN	Southwest	3430 N487WN	71	75	59	19	23	JAX	Jacksonville	BHM	Birmingham	365	3	9	0	N	0	0	0	0	0	19	
19	4	3/1/2019	646	725	655	WN	Southwest	1580 N243WN	99	95	77	30	26	JAX	Jacksonville	BNA	Nashville Int	484	6	16	0	N	0	26	0	4	0	0	
20	4	3/1/2019	1110	1136	1110	WN	Southwest	2195 N479WN	86	90	72	26	30	JAX	Jacksonville	BNA	Nashville Int	484	5	9	0	N	0	0	0	0	10	16	

2. Identify the category of each attribute from the dataset which you have created.

- a. DayOfWeek: Ordinal (1-7 represents the days of the week)
- b. Date: Interval (represents a specific date)
- c. DepTime: Ratio (actual departure time in hours and minutes)
- d. ArrTime: Ratio (actual arrival time in hours and minutes)
- e. CRSArrTime: Ratio (scheduled arrival time in hours and minutes)
- f. UniqueCarrier: Nominal (unique carrier code)
- g. Airline: Nominal (airline company)
- h. FlightNum: Nominal (flight number)

- i. TailNum: Nominal (plane tail number)
- j. ActualElapsedTime: Ratio (actual time spent in the air in minutes)
- k. CRSElapsedTime: Ratio (estimated elapsed time of flight in minutes)
- l. AirTime: Ratio (flight time in minutes)
- m. ArrDelay: Ratio (difference in minutes between scheduled and actual arrival time)
- n. Origin: Nominal (origin IATA airport code)
- o. Org_Airport: Nominal (origin airport name)
- p. Dest: Nominal (destination IATA code)
- q. Dest_Airport: Nominal (destination airport name)
- r. Distance: Ratio (distance between airports in miles)
- s. TaxiIn: Ratio (time taken to taxi from wheels down to arrival at the gate in minutes)
- t. TaxiOut: Ratio (time taken to taxi from departure gate to wheels off in minutes)
- u. Cancelled: Binary (1 = flight canceled, 0 = not canceled)
- v. CancellationCode: Nominal (reason for cancellation)
- w. Diverted: Binary (1 = flight diverted, 0 = not diverted)
- x. CarrierDelay: Ratio (flight delay due to carrier in minutes)
- y. WeatherDelay: Ratio (flight delay due to weather in minutes)
- z. NASDelay: Ratio (flight delay due to National Aviation System in minutes)
- aa. SecurityDelay: Ratio (flight delay due to security reasons in minutes)
- bb. LateAircraftDelay: Ratio (flight delay due to late aircraft arrival in minutes)

3. Identify the attributes which can provide any kind of useful information either collectively or as an individual. Also, discuss the information provided by the attribute and how it will be computed?

Here are some attributes that can provide useful information individually or collectively:

1. **DayOfWeek:** This attribute provides information about the day of the week when the flight occurred. It can be used to analyze patterns in flight schedules, such as peak travel days or days with higher delays. The computation is straightforward as it represents the day of the week on a numerical scale (1-7).
2. **Date:** The date attribute provides information about the specific date of the flight. It can be used to analyze trends over time, such as seasonal variations in flight delays or cancellations. The computation involves representing the date in a suitable format for analysis.
3. **UniqueCarrier:** This attribute represents the unique carrier code for the airline. It provides information about the airline operating the flight. It can be used to analyze the performance of different airlines, compare their on-time records, or identify any specific trends related to particular carriers. The computation involves assigning a unique code to each airline.
4. **ActualElapsedTime:** This attribute represents the actual time spent in the air by the airplane, including taxi-in and taxi-out times. It provides information about the total duration of the flight. It can be used to analyze flight efficiency, compare actual versus scheduled flight times, or identify any significant deviations. The computation involves calculating the difference between the actual departure and arrival times, including taxi times.
5. **Distance:** The distance attribute represents the distance between the origin and destination airports in miles. It provides information about the length of the flight route. It can be used to analyze the relationship between flight distance and other factors like flight time, delays, or fuel consumption. The computation involves calculating the distance between the coordinates of the origin and destination airports.
6. **Cancelled:** This binary attribute indicates whether the flight was canceled or not. It provides information about the flight's operational status. It can be used to analyze the frequency of flight cancellations, identify any patterns or trends, or assess the impact of cancellations on

other attributes like delays or diversion. The computation involves assigning a binary value (1 or 0) based on the flight's cancellation status.

7. CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay: These attributes represent different types of delays and provide information about the reasons for flight delays. Each attribute indicates the delay time caused by a specific factor (carrier, weather, National Aviation System, security, or late aircraft arrival). They can be used to analyze the main causes of delays, identify patterns or trends in delay types, or assess the impact of delays on overall flight operations. The computation involves recording the delay time for each specific factor.

The computation of these attributes involves collecting and recording relevant data during the flight operations, such as departure and arrival times, flight distances, delay reasons, and cancellation status. These attributes collectively provide valuable information for analyzing flight operations, identifying trends, and making informed decisions related to scheduling, performance improvement, and customer satisfaction.

Course outcome:

Understanding of the Data.

Conclusion:

The experiment provided a solid foundation in working with categorical and numerical attributes. We gained practical skills and knowledge that can be applied in various data analysis tasks, such as exploratory data analysis. Understanding the characteristics and nuances of different attribute types is crucial for extracting meaningful insights and making informed decisions in data science and analytics.