

This file is meant to highlights the Excel file that was exported from SQL Server for the flight data exploratory analysis project, as well as the cleaning that was performed to make the file ready for visualization in Tableau.

Note: I cannot publish the Excel file on Github directly, as the file itself is ~125 MB, which is much higher than the allowed 25 MB file limit. However, all steps I performed to clean the data are given here.

After running the query to generate the full joined table in Microsoft SQL Server, I copy and pasted the table into Excel. Below is what the table looked like upon pasting.

	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
	1	origin	time_hour	dep_time	sched_dep_time	del_time	arr_time	sched_arr_time	del_carrier	flight	dest	air_time	distance	day_of_year	day_of_w	month_of_year	net_delay	On Time	Flight_Spe	year	month	day	hour	wind_speed	wind_gust	precip	visib	new_t
2	LGA	00:00:0	839	845	-6	1014	1030	-16	EV	5288	BUF	48	292	10/23/2013	Wednesdi	October	-22	no	365	2013	10	23	8	5.7539	0	0	10	51.
3	EW	00:00:0	840	837	3	959	939	20	EV	4249	BWI	42	169	10/23/2013	Wednesdi	October	23	no	241.429	2013	10	23	8	5.7539	0	0	10	48.
4	JFK	00:00:0	840	845	-5	941	958	-17	9E	3841	PHL	32	94	10/23/2013	Wednesdi	October	-22	no	176.25	2013	10	23	8	5.7539	0	0	10	51.
5	EW	00:00:0	842	848	-6	1116	1110	6	UA	1643	DEN	228	1605	10/23/2013	Wednesdi	October	0	yes	422.368	2013	10	23	8	5.7539	0	0	10	48.
6	JFK	00:00:0	842	850	-8	942	1006	-24	DL	2445	BOS	32	187	10/23/2013	Wednesdi	October	-32	no	350.625	2013	10	23	8	5.7539	0	0	10	51.
7	LGA	00:00:0	843	850	-7	1059	1045	14	MQ	3407	CMH	75	479	10/23/2013	Wednesdi	October	7	yes	383.2	2013	10	23	8	5.7539	0	0	10	51.
8	JFK	00:00:0	843	700	103	1136	1015	81	VX	399	LAX	332	2475	10/23/2013	Wednesdi	October	184	no	447.289	2013	10	23	7	4.60312	0	0	10	48.
9	EW	00:00:0	843	836	7	1056	1046	10	EV	4638	CVG	97	569	10/23/2013	Wednesdi	October	17	yes	351.959	2013	10	23	8	5.7539	0	0	10	48.
10	EW	00:00:0	844	850	-6	1015	1003	12	EV	3810	BUF	52	282	10/23/2013	Wednesdi	October	6	yes	325.385	2013	10	23	8	5.7539	0	0	10	48.
11	LGA	00:00:0	844	850	-6	1009	1028	-19	EV	5286	BTW	36	258	10/23/2013	Wednesdi	October	-25	no	430	2013	10	23	8	5.7539	0	0	10	51.
12	JFK	00:00:0	845	850	-5	1139	1141	-2	B6	189	SAN	333	2446	10/23/2013	Wednesdi	October	-7	yes	440.721	2013	10	23	8	5.7539	0	0	10	51.
13	EW	00:00:0	845	826	19	1055	1014	41	EV	4537	MEM	150	946	10/23/2013	Wednesdi	October	60	no	378.4	2013	10	23	8	5.7539	0	0	10	48.
14	JFK	00:00:0	846	800	46	1025	938	47	B6	885	RDJ	79	427	10/23/2013	Wednesdi	October	93	no	324.304	2013	10	23	8	5.7539	0	0	10	51.
15	EW	00:00:0	847	852	-5	1147	1208	-21	UA	434	SNA	327	2434	10/23/2013	Wednesdi	October	-26	no	446.606	2013	10	23	8	5.7539	0	0	10	48.
16	EW	00:00:0	850	845	5	1202	1140	22	UA	1626	SAN	336	2425	10/23/2013	Wednesdi	October	27	no	433.036	2013	10	23	8	5.7539	0	0	10	48.
17	JFK	00:00:0	850	854	-4	1209	1213	-4	B6	15	SFO	345	2586	10/23/2013	Wednesdi	October	-8	yes	449.739	2013	10	23	8	5.7539	0	0	10	51.
18	EW	00:00:0	851	857	-6	1156	1151	5	UA	997	LAX	317	2454	10/23/2013	Wednesdi	October	-1	yes	464.479	2013	10	23	8	5.7539	0	0	10	48.
19	EW	00:00:0	851	740	71	1120	945	95	EV	4111	CHS	111	628	10/23/2013	Wednesdi	October	166	no	339.459	2013	10	23	8	5.7539	0	0	10	48.
20	LGA	00:00:0	852	855	-3	1159	1156	3	UA	207	IAH	204	1416	10/23/2013	Wednesdi	October	0	yes	416.471	2013	10	23	8	5.7539	0	0	10	51.
21	LGA	00:00:0	852	900	-8	958	1019	-21	US	2140	BOS	33	184	10/23/2013	Wednesdi	October	-29	no	334.545	2013	10	23	9	8.05546	0	0	10	51.
22	EW	00:00:0	852	900	-8	1126	1131	-3	DL	502	ATL	116	746	10/23/2013	Wednesdi	October	-11	yes	385.862	2013	10	23	9	6.90468	0	0	10	48.
23	EW	00:00:0	853	840	13	1224	1145	39	UA	1401	MIA	168	1085	10/23/2013	Wednesdi	October	52	no	387.5	2013	10	23	8	5.7539	0	0	10	48.
24	JFK	00:00:0	853	855	-2	1241	1251	-10	DL	301	SLU	203	1598	10/23/2013	Wednesdi	October	-12	yes	472.315	2013	10	23	8	5.7539	0	0	10	51.
25	EW	00:00:0	855	843	12	1010	957	13	UA	401	BOS	35	200	10/23/2013	Wednesdi	October	25	yes	342.857	2013	10	23	8	5.7539	0	0	10	48.
26	LGA	00:00:0	855	900	-5	1148	1144	4	DL	1747	ATL	120	762	10/23/2013	Wednesdi	October	-1	yes	381	2013	10	23	9	8.05546	0	0	10	51.
27	JFK	00:00:0	855	859	-4	1201	1157	4	B6	795	AUS	221	1521	10/23/2013	Wednesdi	October	0	yes	412.941	2013	10	23	8	5.7539	0	0	10	51.
28	JFK	00:00:0	855	900	-5	1142	1157	-15	B6	354	SEA	320	2422	10/23/2013	Wednesdi	October	-20	yes	454.125	2013	10	23	9	9.20624	0	0	10	51.
29	LGA	00:00:0	856	900	-4	1025	1025	0	US	2173	DCA	46	214	10/23/2013	Wednesdi	October	-4	yes	279.13	2013	10	23	9	8.05546	0	0	10	51.
30	EW	00:00:0	856	900	-4	1224	1208	16	B6	5	FLL	164	1065	10/23/2013	Wednesdi	October	12	no	389.634	2013	10	23	9	6.90468	0	0	10	48.
31	EW	00:00:0	857	835	22	1048	1018	30	EV	4711	STL	140	872	10/23/2013	Wednesdi	October	52	no	373.714	2013	10	23	8	5.7539	0	0	10	48.
32	LGA	00:00:0	858	835	23	1048	1000	48	MQ	3355	BNA	121	764	10/23/2013	Wednesdi	October	71	no	378.843	2013	10	23	8	5.7539	0	0	10	51.
33	EW	00:00:0	858	900	-2	1039	1025	14	UA	1030	ORD	109	719	10/23/2013	Wednesdi	October	12	yes	395.78	2013	10	23	9	6.90468	0	0	10	48.
34	EW	00:00:0	858	900	-2	1212	1206	6	UA	1581	SEA	329	2402	10/23/2013	Wednesdi	October	4	yes	438.055	2013	10	23	9	6.90468	0	0	10	48.
35	JFK	00:00:0	858	900	-2	1205	1211	-6	DL	120	LAX	345	2475	10/23/2013	Wednesdi	October	-8	yes	430.435	2013	10	23	9	9.20624	0	0	10	51.
36	JFK	00:00:0	858	900	-2	1208	1228	-20	DL	431	SFO	343	2586	10/23/2013	Wednesdi	October	-22	no	452.362	2013	10	23	9	9.20624	0	0	10	51.
37	EW	00:00:0	859	902	-3	1206	1148	18	UA	1711	IAH	202	1440	10/23/2013	Wednesdi	October	15	no	415.842	2013	10	23	9	6.90468	0	0	10	48.
38	LGA	00:00:0	901	910	-9	1035	1055	-20	AA	313	ORD	109	733	10/23/2013	Wednesdi	October	-29	no	403.486	2013	10	23	9	8.05546	0	0	10	51.
39	LGA	00:00:0	905	905	0	1109	1115	-6	DL	181	DTW	80	502	10/23/2013	Wednesdi	October	-6	yes	376.5	2013	10	23	9	8.05546	0	0	10	51.
40	LGA	00:00:0	905	830	35	1107	1019	48	EV	5195	RIC	70	292	10/23/2013	Wednesdi	October	83	no	250.286	2013	10	23	8	5.7539	0	0	10	51.
41	JFK	00:00:0	905	910	-5	1207	1220	-13	AA	1	LAX	330	2475	10/23/2013	Wednesdi	October	-18	yes	450	2013	10	23	9	9.20624	0	0	10	51.
42	JFK	00:00:0	905	912	-7	1019	1025	-6	B6	208	PWM	44	273	10/23/2013	Wednesdi	October	-13	yes	372.273	2013	10	23	9	9.20624	0	0	10	51.
43	JFK	00:00:0	907	915	-8	1215	1230	-15	AA	1499	MIA	163	1089	10/23/2013	Wednesdi	October	-23	yes	400.859	2013	10	23	9	9.20624	0	0	10	51.
44	LGA	00:00:0	908	910	-2	1216	1215	1	AA	1121	DFW	201	1389	10/23/2013	Wednesdi	October	-1	yes	414.627	2013	10	23	9	8.05546	0	0	10	51.

Unfortunately, the data is uncleaned, and for visualization in Tableau, I had to clean the data in Excel. What follows is the steps I performed to clean the data.

- 1) I first added a filter to every column of the dataset by **Data -> Sort and Filter -> Filter**
- 2) I froze the top row of the dataset by **View -> Freeze Panes -> Freeze Top Row**
- 3) I switched the data type of the **time_hour** column (Column B) so it displays the correct contents. I did this by first highlighting the entire column and then clicked on the following commands:
Ctrl + 1 -> Number -> Date -> 3/14/12 13:30 -> OK
- 4) Used **Ctrl + H** to replace the NULL values in the **year** column (Column T) to 2013
- 5) Sorted the data by **time_hour** from Oldest to Newest, so that the data is in chronological order
- 6) Used an IFS statement and the **month_of_year** column (Column P) to fill in the missing values
The following is the beginning of the IFS statement in the first row of data in the column:
=IFS(P2='January', 1, P2='February', 2, P2 = 'March', 3, ... P2 = 'December', 12)
- 7) For the columns **visib** (Column AA) and **wind_gust** (Column Y), I used Ctrl + H to replace the blank values with 0
- 8) For the rest of the columns to the right of the **month** column (Column Z through Column AF) I highlighted all of these columns, then:

- a. clicked **Home -> Editing -> Find and Select -> Go to Special -> Blanks -> OK** to only select the blank values
 - b. In the first blank cell, I wrote a formula setting the value of the cell to the one above it
 - c. Clicked **Ctrl + Enter** to apply this to every blank cell
- 9) I deleted out the weather data for 12/31, (from the **hour** column on to the right, so Column W through Column AF, as step 8 filled in this data when no data was given)
- 10) I corrected the **day** column (Column V) for all rows where the date was 12/31/2013 by replacing the 30 in the column with 31
- 11) I filled in the **hour** column for all rows where the date was 12/31 using the following:
- =HOUR(B#)**
- 12) I set the cell width of all columns to 18, so all column names can be read fully
- 13) I bolded and centered the heading of each column by selecting the first row and then clicking **Home -> Font -> B** and **Home -> Alignment -> Center**

After performing the following steps, the dataset now looks like:

	A	B	C	D	E	F	G	H	I	J	K	L	M
	origin	time_hour	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	flight	dest	air_time	distance
2	EWB	1/1/13 5:00	517	515	2	830	819	11 UA		1545 IAH		227	1400
3	LGA	1/1/13 5:00	533	529	4	850	830	20 UA		1714 IAH		227	1416
4	JFK	1/1/13 5:00	542	540	2	923	850	33 AA		1141 MIA		160	1089
5	JFK	1/1/13 5:00	544	545	-1	1004	1022	-18 B6		725 BQN		183	1576
6	EWB	1/1/13 5:00	554	558	-4	740	728	12 UA		1696 ORD		150	719
7	JFK	1/1/13 5:00	559	559	0	702	706	-4 B6		1806 BOS		44	187
8	LGA	1/1/13 6:00	554	600	-6	812	837	-25 DL		461 ATL		116	762
9	EWB	1/1/13 6:00	555	600	-5	913	854	19 B6		507 FLL		158	1065
10	LGA	1/1/13 6:00	557	600	-3	709	723	-14 EV		5708 IAD		53	229
11	JFK	1/1/13 6:00	557	600	-3	838	846	-8 B6		79 MCO		140	944
12	LGA	1/1/13 6:00	558	600	-2	753	745	8 AA		301 ORD		138	733
13	JFK	1/1/13 6:00	558	600	-2	849	851	-2 B6		49 PBI		149	1028
14	JFK	1/1/13 6:00	558	600	-2	853	856	-3 B6		71 TPA		158	1005
15	JFK	1/1/13 6:00	558	600	-2	924	917	7 UA		194 LAX		345	2475
16	EWB	1/1/13 6:00	558	600	-2	923	937	-14 UA		1124 SFO		361	2565
17	LGA	1/1/13 6:00	559	600	-1	941	910	31 AA		707 DFW		257	1389
18	EWB	1/1/13 6:00	559	600	-1	854	902	-8 UA		1187 LAS		337	2227
19	LGA	1/1/13 6:00	600	600	0	851	858	-7 B6		371 FLL		152	1076
20	LGA	1/1/13 6:00	600	600	0	837	825	12 MQ		4650 ATL		134	762
21	EWB	1/1/13 6:00	601	600	1	844	850	-6 B6		343 PBI		147	1023
22	LGA	1/1/13 6:00	602	610	-8	812	820	-8 DL		1919 MSP		170	1020
23	LGA	1/1/13 6:00	602	605	-3	821	805	16 MQ		4401 DTW		105	502
24	EWB	1/1/13 6:00	606	610	-4	858	910	-12 AA		1895 MIA		152	1085
25	JFK	1/1/13 6:00	606	610	-4	837	845	-8 DL		1743 ATL		128	760
26	EWB	1/1/13 6:00	607	607	0	858	915	-17 UA		1077 MIA		157	1085
27	EWB	1/1/13 6:00	608	600	8	807	735	32 MQ		3768 ORD		139	719
28	JFK	1/1/13 6:00	611	600	11	945	931	14 UA		303 SFO		366	2586
29	JFK	1/1/13 6:00	613	610	3	925	921	4 B6		135 RSW		175	1074
30	JFK	1/1/13 6:00	615	615	0	1039	1100	-21 B6		709 SJU		182	1598
31	EWB	1/1/13 6:00	615	615	0	833	842	-9 DL		575 ATL		120	746
32	EWB	1/1/13 6:00	622	630	-8	1017	1014	3 US		245 PHX		342	2133
33	LGA	1/1/13 6:00	623	610	13	920	915	5 AA		1837 MIA		153	1096
34	LGA	1/1/13 6:00	623	627	-4	933	932	1 UA		496 IAH		229	1416
35	EWB	1/1/13 6:00	624	630	-6	909	840	29 EV		4626 MSP		190	1008
36	LGA	1/1/13 6:00	624	630	-6	840	830	10 MQ		4599 MSP		166	1020
37	JFK	1/1/13 6:00	627	630	-3	1018	1018	0 US		27 PHX		330	2153
38	JFK	1/1/13 6:00	628	630	-2	1137	1140	-3 AA		413 SJU		192	1598
39	EWB	1/1/13 6:00	628	630	-2	1016	947	29 UA		1665 LAX		366	2454
40	LGA	1/1/13 6:00	629	630	-1	824	810	14 AA		303 ORD		140	733
41	LGA	1/1/13 6:00	629	630	-1	721	740	-19 WN		4646 BWI		40	185
42	EWB	1/1/13 6:00	629	630	-1	824	833	-9 US		1019 CLT		91	529
43	EWB	1/1/13 6:00	632	608	24	740	728	12 EV		4144 IAD		52	212
44	LGA	1/1/13 6:00	635	635	0	1028	940	48 AA		711 DFW		248	1389

The data is cleaned, looks much more appealing, is more accurate, and is suitable for visualization in Tableau.