

Rzeszów, 03.06.2021

STATYSTYCZNA ANALIZA DANYCH

Projekt

„Wypadki samochodowe”

Witkowski Jakub

Numer indeksu:166733

Inżynieria i Analiza Danych

Grupa L6

1. Opis danych

Dane użyte w projekcie dotyczą wypadków samochodowych w USA w latach 2016-2021 w 49 stanach. Zostały zebrane przez miastowy monitoring, służby porządkowe czujniki drogowe

Dane pochodzą ze strony kaggle.com:

<https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents?resource=download>

Pobrany plik csv został zaimportowany do R i został zapisany do formatu xlsx. Z pliku excel zostały usunięte niektóre kolumny, które nie przydadzą się w projekcie.

Plik zawiera następujące kolumny:

- ID – ID wypadku
- Start_Time – czas, w którym zaczął się wypadek
- End-Time – czas, w którym skończył się wypadek
- Start_Lat – początkowa szerokość geograficzna
- Start_Lng – początkowa długość geograficzna
- End_lat – końcowa szerokość geograficzna
- End_Lng – końcowa długość geograficzna
- Distance.mi – długość wypadku w milach
- Description – krótki opis zdarzenia
- Street – Ulica miejsca zdarzenia
- City – miasto miejsca zdarzenia
- County – powiat miejsca zdarzenia
- State – stan miejsca zdarzenia
- Zipcode – kod pocztowy
- Timezone – strefa czasowa
- Temperature.F – temperatura w Fahrenheit
- Humidity – wilgotność
- Pressure.in – ciśnienie
- Visibility – widoczność
- Wind_direction – kierunek wiatru
- Wind_speed.mph – prędkość wiatru
- Weather_condition – warunki pogodowe

Gotowe dane w excelu:

ID	Start_Time	End_Time	Start_Lat	Start_Lng	End_Lat	End_Lng	Distance.mi.	Description				
A-1	2016-02-08 00:37:08	2016-02-08 06:37:08	40.10891	-83.09286	40.11206	-83.03187	3.23	Between Sawmill Rd/Exit 20 and OH-315/Olentangy Riv Rd/Exit 22 - Accident.				
A-2	2016-02-08 05:36:20	2016-02-08 11:56:20	39.86542	-84.06228	39.86501	-84.04873	0.747	At OH-4/OH-235/Exit 41 - Accident.				
A-3	2016-02-08 06:15:39	2016-02-08 12:15:39	39.10266	-84.52468	39.10209	-84.52396	0.055	At I-71/US-50/Exit 1 - Accident.				
A-4	2016-02-08 06:51:45	2016-02-08 12:51:45	41.06213	-81.53784	41.06217	-81.53547	0.123	At Dart Ave/Exit 21 - Accident.				
A-5	2016-02-08 07:53:43	2016-02-08 13:53:43	39.172393	-84.492792	39.170476	-84.501798	0.5	At Mitchell Ave/Exit 6 - Accident.				
A-6	2016-02-08 08:16:57	2016-02-08 14:16:57	39.06324	-84.03243	39.06731	-84.05851	1.427	At Dela Palma Rd - Accident.				
A-7	2016-02-08 08:15:41	2016-02-08 14:15:41	39.77565	-84.18603	39.77275	-84.18805	0.227	At OH-4/Exit 54 - Accident.				
A-8	2016-02-08 11:51:46	2016-02-08 17:51:46	41.37531	-81.82017	41.36786	-81.82174	0.521	At Bagley Rd/Exit 235 - Accident.				
A-9	2016-02-08 14:19:57	2016-02-08 20:19:57	40.702247	-84.075887	40.69911	-84.084293	0.491	At OH-65/Exit 122 - Accident.				
A-10	2016-02-08 15:16:43	2016-02-08 21:16:43	40.10931	-82.96849	40.11078	-82.984	0.826	At I-71/Exit 26 - Accident.				
A-11	2016-02-08 15:43:50	2016-02-08 21:43:50	39.19288	-84.47723	39.19615	-84.47335	0.307	At OH-4/Paddock Rd/Exit 9 - Accident.				
A-12	2016-02-08 16:50:57	2016-02-08 22:50:57	39.13877	-84.53394	39.13977	-84.5343	0.072	At US-52/Hopple St/Exit 3 - Accident.				
A-13	2016-02-08 17:27:39	2016-02-08 23:27:39	41.47399	-81.704233	41.47386	-81.70559	0.07	At US-42/Exit 170 - Accident.				
A-14	2016-02-08 17:30:18	2016-02-08 23:30:18	39.582242	-83.677814	39.603013	-83.637319	2.59	Between OH-72/Exit 58 and US-35/Exit 65 - Accident.				
A-15	2016-02-08 18:11:11	2016-02-09 00:11:11	40.151785	-81.312635	40.151747	-81.312682	0.004	At Shipley Rd - Accident.				
A-16	2016-02-08 18:11:11	2016-02-09 00:11:11	40.151747	-81.312682	40.151785	-81.312635	0.004	At Titus Rd - Accident.				
A-17	2016-02-08 19:47:42	2016-02-09 01:47:42	39.97241	-82.84695	39.98529	-82.85667	1.028	At OH-16/Broad St/Exit 39 - Accident.				
A-18	2016-02-08 19:47:42	2016-02-09 01:47:42	39.9838	-82.85657	39.9725	-82.84746	0.918	At I-270 - Accident.				
A-19	2016-02-08 20:13:22	2016-02-09 02:13:22	40.02664	-82.9944	40.01228	-82.99218	0.999	Between Weber Rd/Exit 113 and Hudson St/Exit 112 - Accident.				
A-20	2016-02-08 21:00:17	2016-02-09 03:00:17	41.679361	-83.573037	41.666124	-83.566335	0.978	Closed between I-475/Exit 204 and US-24/Detroit Ave/Exit 203 - Road closed due				
A-21	2016-02-08 21:10:10	2016-02-09 03:10:10	40.99613	-85.26613	40.98982	-85.26986	0.477	At Fort Wayne - Accident.				
A-22	2016-02-08 21:30:31	2016-02-09 03:30:31	39.0381	-84.59243	39.03475	-84.60007	0.471	At I-275/Exit 185 - Accident.				
A-23	2016-02-09 05:54:01	2016-02-09 11:54:01	40.03386	-82.99601	40.02246	-82.99515	0.789	Between Broadway/Exit 114 and Weber Rd/Exit 113 - Accident.				
A-24	2016-02-09 06:10:59	2016-02-09 12:10:59	39.85798	-84.28181	39.86441	-84.21554	3.543	Between OH-48/Exit 29 and Dayton Intl Airport Rd/Exit 32 - Accident.				
A-25	2016-02-09 06:46:32	2016-02-09 12:46:32	39.15267	-84.5395	39.14745	-84.53923	0.361	At I-74/US-52/US-27/Exit 4 - Accident.				
A-26	2016-02-09 07:18:10	2016-02-09 13:18:10	39.30732	-85.95982	39.30457	-85.96284	0.249	At US-31/Exit 76 - Accident.				
A-27	2016-02-09 07:28:15	2016-02-09 13:28:15	39.77572	-84.04815	39.76485	-84.09274	2.484	Closed between OH-844/Fairfield Rd/Exit 17 and Colonel Glenn Hwy/Exit 15 - Ro				
A-28	2016-02-09 07:25:40	2016-02-09 13:25:40	39.97643	-83.12306	39.97609	-83.11068	0.656	At I-270/Exit 93 - Accident.				
A-29	2016-02-09 07:44:49	2016-02-09 13:44:49	39.58595	-85.82518	39.58444	-85.82301	0.156	At Fairland Rd/400/Exit 109 - Accident.				
A-30	2016-02-09 07:48:40	2016-02-09 13:48:40	39.3638	-85.51659	39.45279	-85.628	8.554	Between US-421/Exit 132 and US-421/Exit 123 - Accident.				
A-31	2016-02-09 07:48:40	2016-02-09 13:48:40	39.45132	-85.62516	39.36277	-85.51655	8.429	Between US-421/Exit 123 and US-421/Exit 132 - Accident.				
A-32	2016-02-09 08:00:16	2016-02-09 14:00:16	39.75067	-84.14148	39.74883	-84.13402	0.416	At Smithville Rd - Accident.				
A-33	2016-02-09 08:02:55	2016-02-09 14:02:55	39.2675	-81.49929	39.26867	-81.49909	0.082	At US-50/Exit 176 - Accident.				
A-34	2016-02-09 08:18:58	2016-02-09 14:18:58	39.86025	-85.96623	39.86025	-85.96623	0	At High Timber Dr - Accident.				
A-35	2016-02-09 09:07:19	2016-02-09 15:07:19	41.95677	-83.67214	41.95939	-83.67235	0.181	At M-50/Exit 17 - Accident.				
A-36	2016-02-09 10:07:49	2016-02-09 16:07:49	38.27401	-85.74844	38.28558	-85.75518	0.879	Between US-31/IN-62 and Stansifer Ave/Exit 1 - Accident.				
A-37	2016-02-09 11:09:49	2016-02-09 17:09:49	41.95677	-83.67214	41.95939	-83.67235	0.181	At M-50/Exit 17 - Accident.				
A-38	2016-02-09 11:28:49	2016-02-09 17:28:49	40.52225	-80.06666	40.5283	-80.07188	0.5	At Camp Home Rd/Exit 15 - Accident.				
A-39	2016-02-09 12:31:08	2016-02-09 18:31:08	40.48784	-80.00949	40.48614	-80.00908	0.118	At Venture St/Exit 10 - Accident.				
A-40	2016-02-09 15:35:59	2016-02-09 21:35:59	40.156208	-83.018431	40.153321	-83.018205	0.206	Closed at Gold Meadow Dr - Road closed due to accident.				
A-41	2016-02-09 17:12:40	2016-02-09 23:12:40	41.47461	-81.71182	41.47023	-81.74188	1.585	Between 41st St/Exit 169 and OH-10/Lorain Ave/Exit 167 - Accident.				
A-42	2016-02-09 17:13:58	2016-02-09 23:13:58	41.0618	-81.54608	41.06193	-81.54149	0.239	At OH-59/Exit 21 - Accident.				
A-43	2016-02-09 18:20:58	2016-02-10 00:20:58	40.45112	-85.15048	40.35429	-85.14993	6.69	Closed between IN-26 and IN-67 - Road closed due to accident.				
A-44	2016-02-09 18:20:58	2016-02-10 00:20:58	40.35429	-85.14993	40.45112	-85.15048	6.69	Closed between IN-67 and IN-26 - Road closed due to accident.				
A-45	2016-02-09 19:14:08	2016-02-10 01:14:08	40.156208	-83.018431	40.153321	-83.018205	0.206	Closed between IN-67 and IN-26 - Road closed due to accident.				
Street	City	County	State	Zipcode	Timezone	Temperature.F.	Humidity...	Pressure.in.	Visibility.mi.	Wind_Direction	Wind_Speed.mph.	Weather_Condition
Outerbelt E	Dublin	Franklin	OH	43017	US/Eastern	42.1	58	29.76	10 SW		10.4	Light Rain
I-70 E	Dayton	Montgom	OH	45424	US/Eastern	36.9	91	29.68	10 Calm			Light Rain
I-75 S	Cincinnati	Hamilton	OH	45203	US/Eastern	36	97	29.7	10 Calm			Overcast
I-77 N	Akron	Summit	OH	44311	US/Eastern	39	55	29.65	10 Calm			Overcast
I-75 S	Cincinnati	Hamilton	OH	45217	US/Eastern	37	93	29.69	10 WSW		10.4	Light Rain
State Route 32	Williams	Clermont	OH	45176	US/Eastern	35.6	100	29.66	10 WSW		8.1	Overcast
I-75 S	Dayton	Montgom	OH	45404	US/Eastern	33.8	100	29.63	3 SW		2.3	Mostly Cloudy
I-71 S	Cleveland	Cuyahoga	OH	44130	US/Eastern	33.1	92	29.63	0.5 SW		3.5	Snow
E Hanthorn Rd	Lima	Allen	OH	45806	US/Eastern	39	70	29.59	10 WNW		11.5	Overcast
Outerbelt W	Westerville	Franklin	OH	43081	US/Eastern	32	100	29.59	0.5 West		3.5	Snow
I-75 N	Cincinnati	Hamilton	OH	45216	US/Eastern	33.8	100	29.66	3 NNNW		4.6	Light Snow
Hopple St	Cincinnati	Hamilton	OH	45225	US/Eastern	35.1	96	29.69	10 South		3.5	Overcast
I-90 E	Cleveland	Cuyahoga	OH	44113	US/Eastern	33.1	96	29.59	1.8 West		11.5	Light Snow
I-71 N	Jamestown	Greene	OH	45335	US/Eastern	33.8	93	29.64	1 West		5.8	Light Snow
Titus Rd	Freeport	Guernsey	OH	43973	US/Eastern	33.1	92	29.62	10 Calm			Mostly Cloudy
Cadiz Rd	Freeport	Harrison	OH	43973-862	US/Eastern	33.1	92	29.62	10 Calm			Mostly Cloudy
Outerbelt N	Columbus	Franklin	OH	43213	US/Eastern	34	100	29.65	6 Calm			Overcast
Outerbelt S	Columbus	Franklin	OH	43213	US/Eastern	34	100	29.65	6 Calm			Overcast
North Fwy S	Columbus	Franklin	OH	43224	US/Eastern	33.8	100	29.63	1 West		8.1	Light Snow
I-75 S	Toledo	Lucas	OH	43610	US/Eastern	33.4	79	29.62	10 West		8.1	Overcast
US-33 S	Roanoke	Allen	IN	46783	US/Eastern	28	88	29.71	8 West		16.1	Light Snow
I-75 S	Ft Mitche	Kenton	KY	41017	US/Eastern	26.6	80	29.73	3 West		13.8	Light Snow
North Fwy N	Columbus	Franklin	OH	43224	US/Eastern	25	88	29.65	1 SW		9.2	Light Snow
I-70 E	Dayton	Montgom	OH	45415	US/Eastern	23	88	29.66	2 West		15	Light Snow
I-75 S	Cincinnati	Hamilton	OH	45223	US/Eastern	21	88	29.78	1.5 WSW		11.5	Light Snow
I-65 S	Edinburg	Bartholori	IN	46124	US/Eastern	19	100	29.05	2 W		9	Cloudy
I-675 S	Fairborn	Greene	OH	45324	US/Eastern	23	86	29.68	1.5 SW		10.4	Light Snow
I-70 E	Columbus	Franklin	OH	43228	US/Eastern	21.2	93	29.66	1.5 WSW		13.8	Light Snow
US-421 S	Shelbyvill	Shelby	IN	46176	US/Eastern	21.9	82	29.75	1.5 West		12.7	Light Snow
I-74 W	Greensbu	Decatur	IN	47240	US/Eastern	19.4	100	29.78	2 West		9.2	Light Snow
US-421 S	Saint Pau	Decatur	IN	47272	US/Eastern	21.2	86	29.75	3 West		12.7	Light Snow
US-35 E	Dayton	Montgom	OH	45410	US/Eastern	22.5	90	29.73	1.2 SW		9.2	Light Snow
US-50 E	Parkersbu	Wood	WV	26104	US/Eastern	24.1	88	29.72	1 WSW		10.4	Light Snow
Long Lake Ln	Indianap	Marion	IN	46235-612	US/Eastern	19.4	86	29.76	5 WNW		16.1	Overcast
US-23 N	Dundee	Monroe	MI	48131	US/Eastern	30.2	93	29.58	1.2 West		8.1	Light Snow
I-65 S	Jefferson	Clark	IN	47130	US/Eastern	23	74	29.87	10 West		12.7	Overcast
US-23 N	Dundee	Monroe	MI	48131	US/Eastern	31.8	84	29.6	2.5 WNW		12.7	Light Snow
I-279 N	Pittsburg	Allegheni	PA	15237	US/Eastern	30	88	29.56	1.5 West		12.7	Light Snow
US-19 Trk S	Pittsburg	Allegheni	PA	15214	US/Eastern	28.9	82	29.61	1 SW		13.8	Light Snow
Columbus Pike	Lewis Cer	Delaware	OH	43035-941	US/Eastern	26.6	86	29.68	2 West		10.4	Light Snow
Northwest Fwy W	Cleveland	Cuyahoga	OH	44113	US/Eastern	30.9	92	29.62	1 West		19.6	Light Snow
I-76 E	Akron	Summit	OH	44307	US/Eastern	28.9	82	29.62	2.5 SW		9.2	Light Snow
W State Road 26	Dunkirk	Jay	IN	47336	US/Eastern	19.9	81	29.85	2 WNW		12.7	Light Snow
N Meridian St	Redkey	Jay	IN	47373-943	US/Eastern	19.9	81	29.85	2 WNW		12.7	Light Snow
S State Road 1	Ellettsburg	Winnesh	IN	47357-0741	US/Eastern	22	80	29.70	2 WNW		11.5	Light Snow

Rysunek 1 - Dane w excelu

Zaimportowane dane do R Studio

ID	Start_Time	End_Time	Start_Lat	End_Lat	Start_Long	End_Long	State	Description	Street	City	County	State	Zipcode	Temperature_K	Humidity	Pressure_hPa	Visibility_mi	Wind_Speed_mph	Weather_Cond	
1	2016-02-08 08:37:08	2016-02-08 08:37:08	41.10281	41.10280	-83.11296	-83.11317	IL	Between Seawall Rd&St 20 and OH-111&Centerville Rd	Centerville St	Darien	Franklin	OH	43017	US&Canada	42.1	59	29.76	10.00	SNW	10.0 Light Rain
2	2016-02-08 08:38:20	2016-02-08 08:38:20	39.86542	40.04200	-94.06870	-94.06870	IL	At OH-4&OH-23&St 41 - Accident	1702 S	Dayton	Montgomery	OH	45424	US&Canada	38.8	91	29.80	10.00	Caln	At Light Rain
3	2016-02-08 08:39:36	2016-02-08 08:39:36	39.30160	40.13640	-94.13290	-94.13290	IL	At OH-17&St 10&St 1 - Accident	1705 S	Cincinnati	Hamilton	OH	45203	US&Canada	38.0	97	29.70	10.00	Caln	At Overcast
4	2016-02-08 08:40:45	2016-02-08 08:40:45	41.06215	41.15768	-83.15767	-83.15767	IL	At East Ave&St 21 - Accident	177 N	Alton	Summit	OH	44311	US&Canada	38.9	55	29.85	10.00	Caln	At Overcast
5	2016-02-08 08:52:03	2016-02-08 08:52:03	39.37239	40.48270	-94.37048	-94.37048	IL	At Mill&St 4 - Accident	170 S	Cincinnati	Hamilton	OH	45217	US&Canada	37.8	93	29.89	10.00	WNDR	10.0 Light Rain
6	2016-02-08 08:58:17	2016-02-08 08:58:17	39.86549	40.22042	-94.06701	-94.06701	IL	At OH-4&St 16 - Accident	1705 S	Dayton	Montgomery	OH	45424	US&Canada	38.8	90	29.84	10.00	WNDR	2.0 Overcast
7	2016-02-08 09:15:41	2016-02-08 09:15:41	39.75565	40.18061	-94.77070	-94.77070	IL	At OH-4&St 58 - Accident	1705 S	Dayton	Montgomery	OH	45424	US&Canada	38.8	100	29.63	3.00	SNW	2.0 Mostly Cloudy
8	2016-02-08 11:10:48	2016-02-08 11:10:48	41.37031	41.30217	-83.30708	-83.30719	IL	At Maple Rd&St 20 - Accident	172 S	Cleveland	Cuyahoga	OH	44130	US&Canada	33.1	92	29.63	6.50	SNW	3.0 Snow
9	2016-02-08 14:28:17	2016-02-08 08:28:17	40.76255	40.97595	-84.69931	-84.69820	IL	At OH-4&St 122 - Accident	1705 S	Lima	Allen	OH	43009	US&Canada	39.0	70	29.59	10.00	WNDR	11.0 Overcast
10	2016-02-08 15:18:45	2016-02-08 15:18:45	41.30831	42.00040	-82.13070	-82.00040	IL	At OH-17&St 24 - Accident	1705 S	Wentzville	Franklin	OH	43081	US&Canada	32.0	100	29.59	8.50	WNDR	3.0 Snow
11	2016-02-08 15:43:30	2016-02-08 15:43:30	39.30208	40.47702	-94.12015	-94.47702	IL	At OH-4&Puduck Rd&St 9 - Accident	170 N	Cincinnati	Hamilton	OH	45216	US&Canada	37.8	100	29.60	3.00	WNDR	4.0 Light Snow
12	2016-02-08 16:58:17	2016-02-08 16:58:17	39.13877	40.13390	-94.13877	-94.13870	IL	At OH-5&Puduck Rd&St 9 - Accident	1705 S	Cincinnati	Hamilton	OH	45225	US&Canada	38.1	96	29.69	10.00	South	5.0 Overcast
13	2016-02-08 17:27:39	2016-02-08 17:27:39	41.47400	41.76620	-83.47400	-83.76620	IL	At OH-4&St 129 - Accident	170 S	Cincinnati	Cuyahoga	OH	45110	US&Canada	38.1	96	29.59	1.00	WNDR	11.0 Light Snow
14	2016-02-08 17:38:18	2016-02-08 17:38:18	39.56204	41.47702	-94.02001	-83.47702	IL	Between OH-70&St 50 and US-50&St 85 - Accident	171 N	Jeromeville	Greene	OH	45155	US&Canada	33.8	80	29.64	1.00	WNDR	5.0 Light Snow
15	2016-02-08 18:11:33	2016-02-08 18:11:33	41.33178	41.33180	-83.33178	-83.33180	IL	At Shelby Rd - Accident	170 N	Dayton	Montgomery	OH	45373	US&Canada	38.1	92	29.82	10.00	Caln	At Mostly Cloudy
16	2016-02-08 18:18:17	2016-02-08 18:18:17	41.33178	41.33180	-83.33178	-83.33180	IL	At The Rd - Accident	170 N	Dayton	Montgomery	OH	45373	US&Canada	38.1	92	29.82	10.00	Caln	At Mostly Cloudy
17	2016-02-08 18:47:42	2016-02-08 18:47:42	39.97461	42.04400	-94.96920	-84.04400	IL	At OH-16&St 10 - Accident	170 N	Columbus	Franklin	OH	43215	US&Canada	38.9	100	29.85	6.00	Caln	At Overcast
18	2016-02-08 18:47:42	2016-02-08 18:47:42	39.80300	42.04400	-94.96920	-84.04400	IL	At OH-16 - Accident	170 N	Columbus	Franklin	OH	43215	US&Canada	38.8	100	29.85	6.00	Caln	At Overcast
19	2016-02-08 20:12:22	2016-02-08 20:12:22	42.04400	42.04400	-94.96920	-84.04400	IL	Between Weber Rd&St 112 and Weber Rd&St 112 - Accident	Northway S	Columbus	Franklin	OH	43224	US&Canada	38.8	100	29.85	2.00	WNDR	8.0 Light Snow
20	2016-02-08 21:08:17	2016-02-08 21:08:17	41.17636	41.17636	-83.18613	-83.18613	IL	Closed between 14 Tufford Rd and US-36&St 16 - Accident	170 S	Nobles	Lucas	OH	43010	US&Canada	33.4	79	29.83	10.00	WNDR	8.0 Overcast
21	2016-02-08 21:18:18	2016-02-08 21:18:18	41.06613	41.26613	-83.26613	-83.26613	IL	At Fort Wayne - Accident	US-33 S	Franklin	Allen	IN	46103	US&Canada	38.0	80	29.71	6.00	WNDR	10.0 Light Snow
22	2016-02-08 21:38:18	2016-02-08 21:38:18	39.83816	40.06816	-94.06816	-94.06816	IL	At OH-17&St 200 - Accident	170 S	Ellettsville	Allen	IN	46103	US&Canada	38.0	80	29.71	6.00	WNDR	11.0 Light Snow
23	2016-02-08 22:38:18	2016-02-08 22:38:18	41.03000	42.03000	-82.03000	-82.03000	IL	Between Broad&St 114 and Weber Rd&St 112 - Accident	Northway S	Columbus	Franklin	OH	43224	US&Canada	38.8	80	29.85	1.00	SNW	9.0 Light Snow
24	2016-02-08 22:38:18	2016-02-08 22:38:18	39.87708	40.21010	-94.06841	-84.21010	IL	Between OH-4&St 26 and Dayton Int Airport Rd - Accident	170 S	Dayton	Montgomery	OH	45415	US&Canada	37.8	80	29.64	2.00	WNDR	13.0 Light Snow
25	2016-02-08 22:48:18	2016-02-08 22:48:18	39.13877	40.13390	-94.13877	-94.13390	IL	At OH-5&St 10 - Accident	170 S	Cincinnati	Hamilton	OH	45225	US&Canada	37.8	100	29.69	1.00	WNDR	13.0 Light Snow
26	2016-02-08 23:18:18	2016-02-08 23:18:18	39.30710	41.03000	-94.30407	-84.03000	IL	At OH-16&St 10 - Accident	140 S	Dayton	Montgomery	OH	45424	US&Canada	38.8	100	29.85	2.00	SNW	9.0 Cloudy
27	2016-02-08 23:28:18	2016-02-08 23:28:18	39.77570	40.06816	-94.06816	-94.06816	IL	Between OH-16&St 10 and OH-16&St 10 - Accident	140 S	Dayton	Montgomery	OH	45424	US&Canada	38.8	100	29.85	2.00	SNW	10.0 Light Snow
28	2016-02-08 23:28:18	2016-02-08 23:28:18	39.87461	42.12000	-94.97000	-84.12000	IL	At OH-16&St 10 - Accident	170 S	Columbus	Franklin	OH	43228	US&Canada	37.2	93	29.86	1.50	WNDR	13.0 Light Snow
29	2016-02-08 23:48:18	2016-02-08 23:48:18	39.50500	41.03000	-94.03000	-84.03000	IL	At National Road&St 100 - Accident	US-41 S	Jeffersville	Shelby	IN	46174	US&Canada	38.8	80	29.70	1.50	WNDR	12.0 Light Snow
30	2016-02-08 23:48:18	2016-02-08 23:48:18	39.36300	41.11010	-94.42000	-84.42000	IL	Between US-41&St 102 and US-41&St 120 - Accident	174 W	Greenburg	Decatur	IN	47040	US&Canada	38.4	100	29.70	2.00	WNDR	9.0 Light Snow
31	2016-02-08 23:48:18	2016-02-08 23:48:18	39.41310	41.42010	-94.36270	-84.42010	IL	Between US-41&St 102 and US-41&St 120 - Accident	US-41 S	Saint Paul	Decatur	IN	47272	US&Canada	37.2	96	29.70	3.00	WNDR	12.0 Light Snow
32	2016-02-08 23:48:18	2016-02-08 23:48:18	39.70000	40.14140	-94.70000	-84.14140	IL	At Smith&St 10 - Accident	US-41 S	Dayton	Montgomery	OH	45420	US&Canada	37.2	90	29.71	1.00	SNW	9.0 Light Snow
33	2016-02-08 23:48:18	2016-02-08 23:48:18	39.36700	41.49000	-94.36700	-84.49000	IL	At US-50&St 176 - Accident	US-50 S	Hickoryburg	Way	IN	46104	US&Canada	38.1	96	29.72	1.00	WNDR	10.0 Light Snow
34	2016-02-08 23:48:18	2016-02-08 23:48:18	39.86025	41.06025	-94.06025	-84.06025	IL	At High Timber Dr - Accident	Long Lake Ln	Indianapolis	Marion	IN	46220-0127	US&Canada	39.4	96	29.76	5.00	WNDR	10.0 Overcast
35	2016-02-08 23:48:18	2016-02-08 23:48:18	41.00000	41.00000	-83.00000	-83.00000	IL	At OH-16&St 17 - Accident	US-17 S	Dayton	Montgomery	OH	45420	US&Canada	38.2	93	29.50	1.00	SNW	9.0 Light Snow

Rysunek 2 - Dane w R Studio

2. Podstawowe parametry opisowe

W projekcie zostały wyznaczone: średnia, mediana, odchylenie standardowe, kwartyle, rozstęp międzykwartyłowy, wariancja, moment centralny, przedział zmienności, kurtoza i współczynnik skośności. Parametry zostały policzone dla długości wypadku podanej w milach oraz dla ilości wypadków w latach 2016-2021.

Średnia:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Mediana – środkowa wartość w szeregu uporządkowanym rosnącym.

Odchylenie standardowe - mówi, jak szeroko wartości jakiejś są rozrzucone wokół jej średniej.

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 n_i - (\bar{x})^2}$$

Rozstęp międzykwartyłowy - różnica między trzecim a pierwszym kwartylem.

Wariancja - średnia arytmetyczna kwadratów odchyleń od ich średniej arytmetycznej.

$$s = \frac{1}{n} \sum_{i=1}^n x_i^2 n_i - (\bar{x})^2$$

Moment centralny - średnia arytmetyczna z odchyleń poszczególnych wartości zmiennej od średniej arytmetycznej podniesionych do r-tej potęgi.

Kurtoza – miara koncentracji i spłaszczenia rozkładu. Określa rozmieszczenie i koncentrację wartości w pobliżu średniej.

Współczynnik skośności - służy do określania jak wygląda rozkład, tzn. czy dane są w miarę równo rozłożone po obu stronach średniej/mediany, czy może po którejś stronie obserwacje leżą dalej od średniej.

Powyższe parametry w R Studio:

	Parametr	Wynik
1	Średnia	0.8053732147269100360987
2	Mediana	0.3360000000000000208722
3	Odchylenie standardowe	1.5266983956037574188258
4	Kwartył 0%	0.00000000000000000000
5	25%	0.106999999999999978906
6	50%	0.3360000000000000208722
7	75%	0.8890000000000000124345
8	100%	151.5250000000000056843419
9	Rozstęp międzykwartyłowy	0.7820000000000000284217
10	Wariancja	2.3308079911390873206756
11	Moment centralny	-0.0000000000000000475637
12	Przedział zmienności Min	0.00000000000000000000
13	Max	151.5250000000000056843419
14	Kurtoza	603.3679747449089063593419
15	Współczynnik skośności	12.9956858035041484811245

Rysunek 3 - Ramka danych z parametrami opisowymi dla długości wypadku

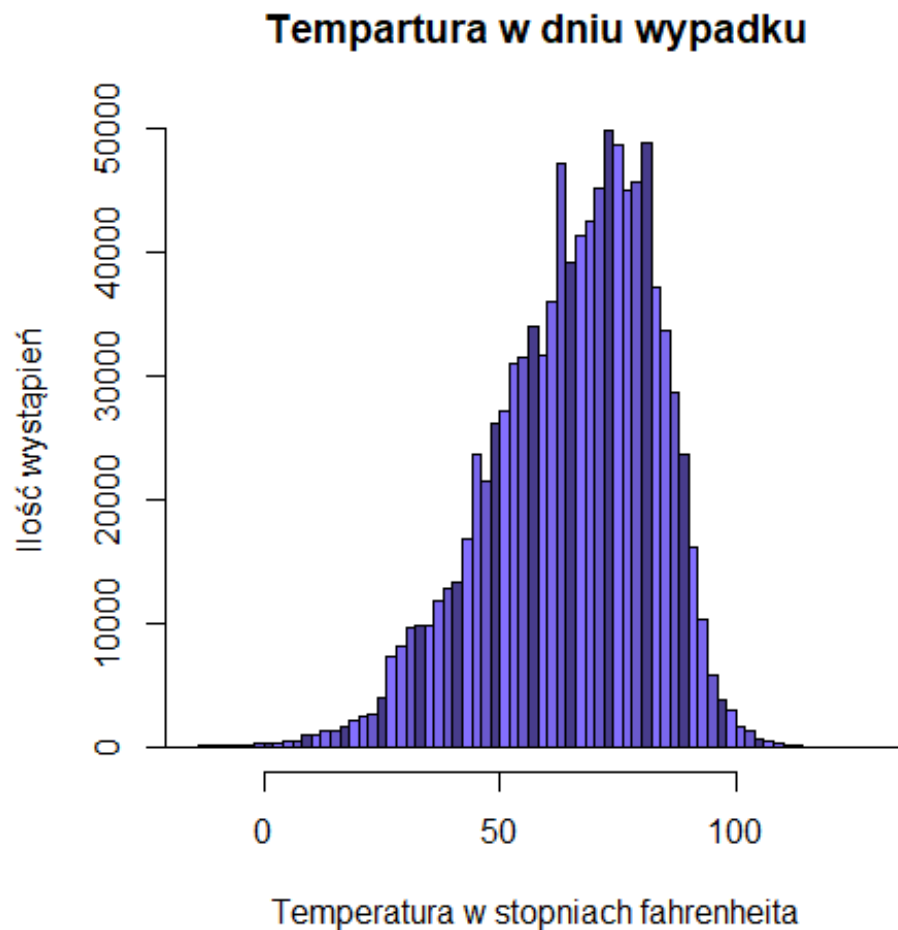
	Parametr2	Wynik2
1	Średnia	170667.16666666665696539
2	Mediana	51483.500000000000000000
3	Odchylenie standardowe	312773.36215439875377342
4	Kwartył 0%	1.00000000000000000000
5	25%	12.25000000000000000000
6	50%	51483.500000000000000000
7	75%	117248.250000000000000000
8	100%	799010.000000000000000000
9	Rozstęp międzykwartyłowy	117236.000000000000000000
10	Wariancja	97827176073.36666870117187500
11	Moment centralny	0.00000000001455192
12	Przedział zmienności Min	1.00000000000000000000
13	Max	799010.000000000000000000
14	Kurtoza	3.97359185433802109
15	Współczynnik skośności	1.66651703492831982

Rysunek 4 - Ramka danych z parametrami opisowymi dla ilości wypadków

3. Graficzna prezentacja danych

Histogram

Wykres przedstawia histogram dla temperatur dla każdego wypadku z ramki danych.

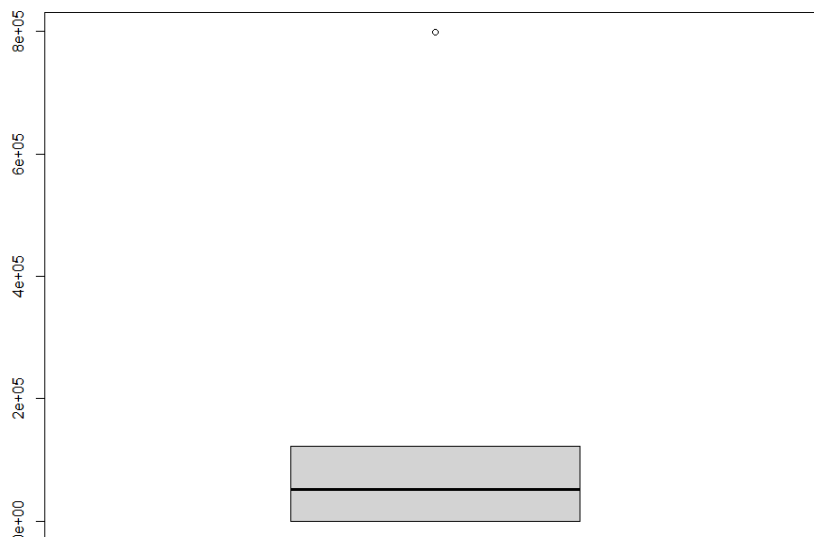


Rysunek 5 - Histogram temperatur

Oś pozioma pokazuje stopnie w skali Fahrenheita, natomiast pionowa ilość wystąpień danej temperatury.

Wykres pudełkowy

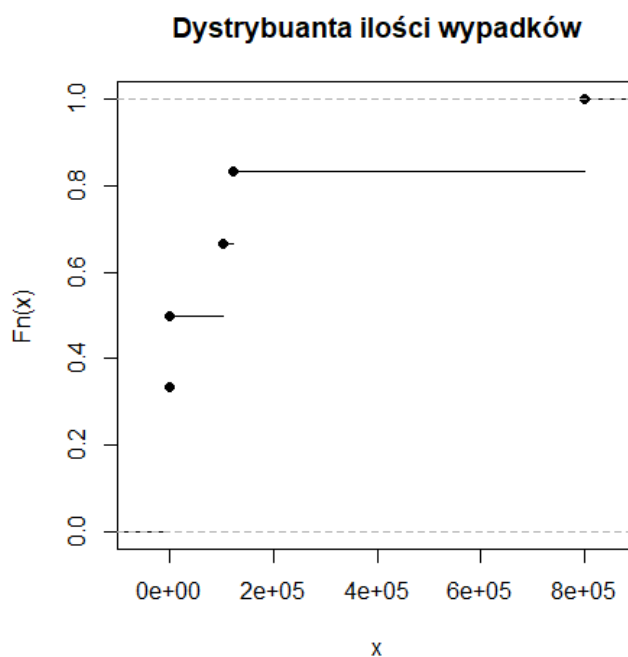
Wykres przedstawia ilość wypadków dla każdego roku. W tym celu została stworzona nowa ramka danych tylko z kolumną Start_Time. Następnie za pomocą funkcji separate kolumna Start_Time została podzielona na 4 osobne kolumny: rok, dzień, miesiąc oraz czas. Następnie policzona została ilość wystąpień danych lat: 2016, 2017, 2018, 2019, 2020, 2021. Została stworzona nowa ramka danych ilość_wypadkow_w_roku. Do wykresu pudełkowego została użyta funkcja boxplot().



Rysunek 6 - Wykres pudełkowy

Wykres dystrybuanty

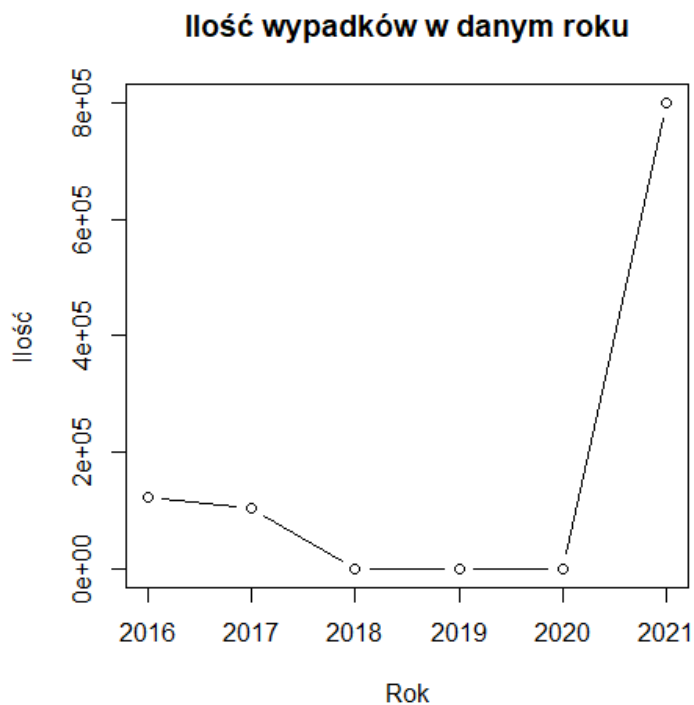
Na wykresie oś pionowa pokazuje zakres wartości dystrybuanty od 0 do 1, a oś pozioma wyniki zmiennej.



Rysunek 7 - Wykres dystrybuanty

Wykres liniowy

Wykres przedstawia ilość wypadków w latach 2016-2021. Na osi poziomej pokazane są dane lata, natomiast na pionowej ilość wypadków.

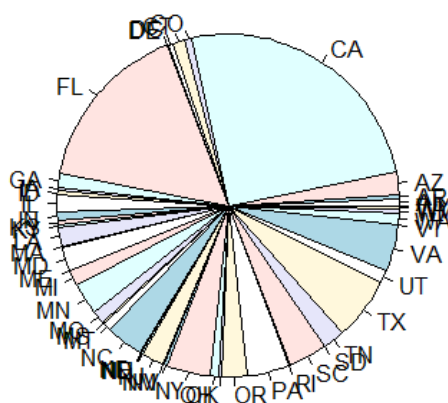


Rysunek 8 - Wykres liniowy

Wykres kołowy

Wykres kołowy przedstawia, w jakim stanie jest najwięcej wypadków.

Wykres kołowy wypadków dla każdego stanu



Rysunek 9 - Wykres kołowy

4. Weryfikacja hipotez statystycznych

Pierwszą hipotezą jest weryfikacja wartości średniej dla ilości wypadków w roku.

Dla hipotezy zerowej wartość średniej wynosi 100 000.

$$H_0: \mu = 100000$$

Hipotezą alternatywną będzie wartość różna od 100 000.

$$H_1: \mu \neq 100000$$

Aby sprawdzić hipotezę należy użyć funkcji t.test:

```
> t.test(x, mu=500000)

      One Sample t-test

data:  x
t = -2.5792, df = 5, p-value = 0.04948
alternative hypothesis: true mean is not equal to 5e+05
95 percent confidence interval:
 -157568.3  498902.7
sample estimates:
mean of x
 170667.2
```

Rysunek 10 - Wynik pierwszej hipotezy

W zmiennej x jest przechowywana ilość wypadków w latach 2016-2021. Na podstawie podanego wyniku, możemy stwierdzić, że $p < 0.05$, więc mamy podstawę do odrzucenia hipotezy zerowej. Z otrzymanego komunikatu można również odczytać przedział ufności wynoszący 95% oraz wartość średniej ze zmiennej x.

Drugą hipotezą jest weryfikacja wartości średniej dla temperatury.

Dla hipotezy zerowej mamy:

$$H_0: \mu = 20$$

Dla hipotezy alternatywnej:

$$H_1: \mu \neq 20$$

Wynik funkcji:

```
> t.test(y, mu=20)

      One Sample t-test

data:  y
t = 2640.2, df = 1002758, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 20
95 percent confidence interval:
  65.74282  65.81079
sample estimates:
mean of x
 65.7768
```

Rysunek 11 - Wynik drugiej hipotezy

W zmiennej *y* jest przechowywana wartość temperatur w dniu każdego wypadku. Wartość *p* jest mniejsza niż $2.2e-16$, więc jest to dużo mniej niż 0.05. Mamy więc podstawę do odrzucenia hipotezy zerowej. Z otrzymanego komunikatu można również odczytać przedział ufności wynoszący 95% oraz wartość średniej.

5. Wnioski

Celem projektu było wyznaczenie parametrów opisowych oraz sprawdzenie hipotez statystycznych. W projekcie zostały policzone parametry takie jak średnia, mediana, odchylenie standardowe, kwantyle. Zostały użyte różnego rodzaju wykresy tj. Histogram, wykres dystrybucyjny, kołowy, liniowy. Została pogłębiona wiedza w weryfikacji hipotez. W trakcie analizowania wypadków można było się dowiedzieć, że najwięcej wypadków było w stanach takich jak Kalifornia, Floryda czy Teksas, oraz, że na wypadki składa się wiele czynników takich jak widoczność na drodze, czy aktualna pogoda. Projekt pozwolił mi na rozwinięcie umiejętności w analizowaniu danych oraz w programowaniu w R Studio.

6. Lista komend

Parametry opisowe:

```
średnia_dlugosci_wypadku<-mean(dane$Distance.mi.)
mediana_dlugosci_wypadku<-median(dane$Distance.mi.)
odch_stand_dlugosci_wypadku<-sd(dane$Distance.mi., na.rm=TRUE)
kwartyl_dlugosci_wypadku<-quantile(dane$Distance.mi.)
rozstep_miedzykwartylowy_dlugosci_wypadku<-IQR(dane$Distance.mi.)
wariancja_dlugosci_wypadku<-var(dane$Distance.mi.)
moment_cent_dlugosci_wypadku<-moment(dane$Distance.mi., central=TRUE)
przedzial_zmiennosci_dlugosci_wypadku<-range(dane$Distance.mi.)
kurtoza_dlugosci_wypadku<-kurtosis(dane$Distance.mi.)
wsp_skos_dlugosci_wypadku<-skewness(dane$Distance.mi.)
srednia_ilosci_wypadkow<-mean(ilosc_wypadkow_w_roku$Ilość)
mediana_ilosci_wypadkow<-median(ilosc_wypadkow_w_roku$Ilość)
odch_stand_ilosci_wypadkow<-sd(ilosc_wypadkow_w_roku$Ilość, na.rm=TRUE)
kwartyl_ilosci_wypadkow<-quantile(ilosc_wypadkow_w_roku$Ilość)
rozstep_miedzykwartylowy_ilosci_wypadkow<-IQR(ilosc_wypadkow_w_roku$Ilość)
wariancja_ilosci_wypadkow<-var(ilosc_wypadkow_w_roku$Ilość)
moment_cent_ilosci_wypadkow<-moment(ilosc_wypadkow_w_roku$Ilość,
central=TRUE)
```

```
przedzial_zmiennosci_ilosci_wypadkow<-range(ilosc_wypadkow_w_roku$Ilość)
```

```
kurtoza_ilosci_wypadkow<-kurtosis(ilosc_wypadkow_w_roku$Ilość)
```

```
wsp_skos_ilosci_wypadkow<-skewness(ilosc_wypadkow_w_roku$Ilość)
```

Histogram

```
hist(dane$Temperature.F., main="Tempartura w dniu wypadku", xlab="Temperatura  
w stopniach fahrenheit", ylab="Ilość wystąpień",breaks = 100,xlim=c(-  
15,130),col=c("slateblue1", "slateblue2", "slateblue3", "slateblue4"))
```

Wykres pudełkowy

```
ggplot(ilosc_wypadkow_w_roku, aes(x=Rok, y=Ilość,  
group=1))+geom_boxplot()+coord_flip()  
  
boxplot(ilosc_wypadkow_w_roku$Ilość)
```

Wykres dystrybuanty

```
plot(ecdf(ilosc_wypadkow_w_roku$Ilość), main="Dystrybuanta ilości wypadków")
```

Wykres liniowy

```
plot(ilosc_wypadkow_w_roku$Rok,ilosc_wypadkow_w_roku$Ilość, type="b",  
main="Ilość wypadków w danym roku", xlab="Rok", ylab="Ilość")
```

Wykres kołowy

```
pie(Stan$Freq, Stan$stany, main="Wykres kołowy wypadków dla każdego stanu")
```

Weryfikacja hipotez

```
x<-ilosc_wypadkow_w_roku$Ilość
```

```
t.test(x, mu=500000)
```

```
y<-dane$Temperature.F.
```

```
t.test(y, mu=20)
```

7. Użyte funkcje

- Read.csv – funkcja do wczytywania plików z rozszerzeniem csv, pakiet wbudowany w R studio
- Write_xlsx – funkcja do zapisywania plików Excel, pochodzi z pakietu writexl
- Read_excel - funkcja do wczytywania plików z rozszerzeniem xlsx, pochodzi z pakietu readxl
- Mean() – średnia, pakiet wbudowany w R studio
- Median() – mediana, pakiet stats
- Sd() – odchylenie standardowe, pakiet stats
- Quantile() – kwantyle, pakiet stats
- IQR() – rozstęp międzykwartyłowy, pakiet stats
- Var() – wariancja, pakiet stats
- Moment() – moment centralny, pakiet moments

- Range() – przedział zmienności, pakiet wbudowany
- Kurtosis() – kurtoza, pakiet moments
- Skewness() – współczynnik skośności, pakiet moments
- Data.frame() – funkcja do tworzenia ramki danych, pakiet wbudowany
- Format(x, scientific=FALSE) – argument funkcji, który zmienia zapis danych z np. 8.053732e-01 na 0.8053732147269100360987, pakiet wbudowany
- Hist() – funkcja tworząca histogram, pakiet graphics
- Ggplot()+geom_boxplot(), boxplot() – funkcje tworzące wykres pudełkowy, pakiet tidyverse
- Plot() – funkcja tworząca wykres liniowy, pakiet wbudowany
- Pie() – funkcja tworząca wykres kołowy, pakiet graphics
- T.test() – funkcja do sprawdzania hipotez, pakiet stats

8. Skrypt projektu w R Studio

```
#Statystyczna analiza danych
```

```
#Witkowski Jakub
```

```
#166733
```

```
#ładowanie potrzebnych pakietów
```

```
library(readxl)
```

```
library(dplyr)
```

```
library(writexl)
```

```
#install.packages("moments")
```

```
library(moments)
```

```
#install.packages("data.table")
```

```
library(data.table)
```

```
library(tidyr)
```

```
library(tidyverse)
```

```
#załadowanie danych do ramki danych
```

```
dane<-(read_excel("C:/Users/Kuba/Desktop/studia/Semestr 4/Statystyczna analiza  
danych/Projekt/dane.xlsx"))
```

```
#obliczanie parametrów opisowych
```

```
średnia_dlugosci_wypadku<-mean(dane$Distance.mi.) #średnia
```

```

mediana_dlugosci_wypadku<-median(dane$Distance.mi.) #mediana

odch_stand_dlugosci_wypadku<-sd(dane$Distance.mi., na.rm=TRUE) #odchylenie
standardowe

kwartyl_dlugosci_wypadku<-quantile(dane$Distance.mi.) #kwantyle

rozstep_miedzykwartylowy_dlugosci_wypadku<-IQR(dane$Distance.mi.) #rozstęp
miedzykwartylowy

wariancja_dlugosci_wypadku<-var(dane$Distance.mi.) #wariancja

moment_cent_dlugosci_wypadku<-moment(dane$Distance.mi., central=TRUE)
#moment centralny

przedzial_zmienności_dlugosci_wypadku<-range(dane$Distance.mi.) #przedziały
zmienności

kurtoza_dlugosci_wypadku<-kurtosis(dane$Distance.mi.) #kurtoza

wsp_skos_dlugosci_wypadku<-skewness(dane$Distance.mi.) #współczynnik skośności


#tworzenie 2 kolumn do ramki danych
Parametr<-c("Średnia", "Mediana", "Odchylenie standardowe",
            "Kwartyl 0%", "25%", "50%", "75%", "100%", "Rozstęp miedzykwartylowy",
            "Wariancja",
            "Moment centralny", "Przedział zmienności Min", "Max", "Kurtoza",
            "Współczynnik skośności")

Wynik<-c(średnia_dlugosci_wypadku, mediana_dlugosci_wypadku,
        odch_stand_dlugosci_wypadku, kwartyl_dlugosci_wypadku,
        rozstep_miedzykwartylowy_dlugosci_wypadku, wariancja_dlugosci_wypadku,
        moment_cent_dlugosci_wypadku, przedzial_zmienności_dlugosci_wypadku,
        kurtoza_dlugosci_wypadku, wsp_skos_dlugosci_wypadku)

#ramka danych zawierające parametry opisowe
parametry_opisowe_dlugosci_wypadku<-data.frame(Parametr, Wynik=format(Wynik,
scientific=FALSE))


#tworzenie histogramu temperatur

hist(dane$Temperature.F., main="Tempartura w dniu wypadku", xlab="Temperatura
w stopniach fahrenheit",
     ylab="Ilość wystąpień", breaks = 100, xlim=c(-15,130), col=c("slateblue1",
"slateblue2", "slateblue3", "slateblue4"))

```

```

df<-data.frame(dane$Start_Time)
#rozdzielanie kolumny na pojedyncze kolumny
df<-separate(df, dane.Start_Time, c("rok", "dzień", "miesiąc", "czas"), "-")
#zliczanie ilości wystąpień w kolumnie
lata<-df$rok
lata<-table(lata)
a<-lata[names(lata)==2016]
b<-lata[names(lata)==2017]
c<-lata[names(lata)==2018]
d<-lata[names(lata)==2019]
e<-lata[names(lata)==2020]
f<-lata[names(lata)==2021]
Rok<-c("2016", "2017", "2018", "2019", "2020", "2021")
Ilość<-c(a,b,c,d,e,f)
ilosc_wypadkow_w_roku<-data.frame(Rok, Ilość)

#wykres pudełkowy ilości wypadków w danych latach
ggplot(ilosc_wypadkow_w_roku, aes(x=Rok, y=Ilość,
group=1))+geom_boxplot()+coord_flip()
boxplot(ilosc_wypadkow_w_roku$Ilość)

#wykres dystrybucyjny
plot(ecdf(ilosc_wypadkow_w_roku$Ilość), main="Dystrybuanta ilości wypadków")

#wykres liniowy
plot(ilosc_wypadkow_w_roku$Rok,ilosc_wypadkow_w_roku$Ilość, type="b",
      main="Ilość wypadków w danym roku", xlab="Rok", ylab="Ilość")

#wykres kołowy
stany<-dane$State
Stan<-as.data.frame(table(stany))

```

```
pie(Stan$Freq, Stan$stany, main="Wykres kołowy wypadków dla każdego stanu")
```

```
#parametry opisowe ilości wypadków
```

```
srednia_ilosci_wypadkow<-mean(ilosc_wypadkow_w_roku$Ilość)
```

```
mediana_ilosci_wypadkow<-median(ilosc_wypadkow_w_roku$Ilość)
```

```
odch_stand_ilosci_wypadkow<-sd(ilosc_wypadkow_w_roku$Ilość, na.rm=TRUE)
```

```
kwartyl_ilosci_wypadkow<-quantile(ilosc_wypadkow_w_roku$Ilość)
```

```
rozstep_miedzykwartylowy_ilosci_wypadkow<-IQR(ilosc_wypadkow_w_roku$Ilość)
```

```
wariancja_ilosci_wypadkow<-var(ilosc_wypadkow_w_roku$Ilość)
```

```
moment_cent_ilosci_wypadkow<-moment(ilosc_wypadkow_w_roku$Ilość,  
central=TRUE)
```

```
przedzial_zmiennosci_ilosci_wypadkow<-range(ilosc_wypadkow_w_roku$Ilość)
```

```
kurtoza_ilosci_wypadkow<-kurtosis(ilosc_wypadkow_w_roku$Ilość)
```

```
wsp_skos_ilosci_wypadkow<-skewness(ilosc_wypadkow_w_roku$Ilość)
```

```
Parametr2<-c("Średnia", "Mediana", "Odchylenie standardowe",
```

```
          "Kwartyl 0%", "25%", "50%", "75%", "100%", "Rozstęp międzykwartylowy",  
          "Wariancja",
```

```
          "Moment centralny", "Przedział zmienności Min", "Max", "Kurtoza",  
          "Współczynnik skośności")
```

```
Wynik2<-c(srednia_ilosci_wypadkow, mediana_ilosci_wypadkow,  
odch_stand_ilosci_wypadkow,
```

```
          kwartyl_ilosci_wypadkow, rozstep_miedzykwartylowy_ilosci_wypadkow,
```

```
          wariancja_ilosci_wypadkow, moment_cent_ilosci_wypadkow,  
przedzial_zmiennosci_ilosci_wypadkow,
```

```
          kurtoza_ilosci_wypadkow, wsp_skos_ilosci_wypadkow)
```

```
parametry_opisowe_ilosci_wypadkow<-data.frame(Parametr2,  
Wynik2=format(Wynik2, scientific=FALSE))
```

```
#Weryfikacja hipotez statystycznych
```

```
x<-ilosc_wypadkow_w_roku$Ilość
```

```
mean(x)
```

```
t.test(x, mu=500000)
```

```
y<-dane$Temperature.F.
```

```
mean(y)
```

```
t.test(y, mu=20)
```