

Report

Data description:

This dataset reports statistics for live births occurring within the United States to U.S. residents. Data are available by a variety of demographic characteristics, such as state and county of residence, mother's race, and mother's age, and health and medical items, such as tobacco use, method of delivery, and congenital anomalies. The data are derived from birth certificates issued in years 2016-2020.

id	birth_year	birth_month	...	plurality	gender	infant_weight
1	2016	1	...	1	F	3705
2	2016	1	...	1	F	3771
4	2016	1	...	1	M	3459
5	2016	1	...	1	M	3320
6	2016	1	...	1	M	4300

Features in data:

Id - A unique number (1-180992).

Birth_year - What year did the birth occur (2016-2020).

Birth_month - What month did the birth occur (1-12).

Birth_time - What time did the birth occur (0000 - 2359).

Birth_place – Where the birth took place (1- Hospital, 2- Freestanding Birth Center, 3 - Clinic/ Doctor's Office, 4- Residence,5- Other,6- Unknown or Not Stated ,'' - Unknown).

Mother_age - The age of the mother at the time of birth (13-50).

Marital_status - The marital status of the mother as recorded on the birth certificate(1- Married,2- Unmarried , ' ' - Unknown) .

Mother_education - This field indicates a range for the number of years of education received by the mother at the time of birth (1- 8th grade or less ,2- 9th through 12th grade with no diploma,3- High school graduate or GED completed ,4 - Some college credit, but not a degree , 5- Associate degree (AA, AS) , 6- Bachelor's degree (BA, AB, BS) , 7- Master's degree (MA, MS) , 8 - Doctorate (PHD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD) , ' ' - Unknown).

Father_age - The age of the father at the time of birth (13-91, `` - Unknown).

Father education - This field indicates a range for the number of years of education received by the father at the time of birth (1- 8th grade or less ,2- 9th through 12th grade with no diploma,3- High school graduate or GED completed ,4 - Some college credit, but not a degree , 5- Associate degree (AA, AS) , 6- Bachelor's degree (BA, AB, BS) , 7- Master's degree (MA, MS) , 8 - Doctorate (PHD, EdD) or Professional Degree (MD, DDS, DVM, LLB, JD) , ' ' - Unknown).

Interval llb - This field reports the time elapsed since the mother's last live birth in months,two columns are provided, a column for the code value and a column for the label.(3-300,888- Not Applicable (1st live birth) , ' ' - Unknown).

Cigarettes - Number of Cigarettes Before Pregnancy (0-97,98 or more, ' ' - Unknown).

Mother height - Mother's Height in Inches (30 -78, ' ' - Unknown).

Mother bmi - Calculates the average pre-pregnancy Body Mass Index (BMI) value for mothers (13-69.5, ' ' - Unknown)

Pre preg weight – Mother's pre-pregnancy weight in lbs (75-375, ' ' - Unknown)

Delivery weight – Mother's delivery weight in lbs (100-400, ' ' - Unknown)

Pre preg diabetes - Pre-pregnancy diabetes that woman have diabetes before you get pregnant (1- Y , 2 – N, ' ' - Unknown).

Gest diabetes - Gestational Diabetes that women get during pregnancy (1- Y , 2 – N, ' ' - Unknown).

Pre preg hypertension - Prepregnancy Hypertension reports a diagnosis prior to the onset of this pregnancy, and does not include gestational (1- Y , 2 – N, ' ' - Unknown).

Gest hypertension - Gestational Hypertension reports whether a diagnosis occurred in this pregnancy of pregnancy-induced Hypertension or preeclampsia (1- Y , 2 – N, ' ' - Unknown).

Prev preterm birth- This field indicates whether previous preterm birth is reported as a maternal risk factor (1- Y , 2 – N, ' ' - Unknown).

Infertility treatment - Yes is reported when pregnancy resulted from infertility treatment including fertility-enhancing drugs; artificial insemination or intrauterine insemination (1- Y , 2 – N, ' ' - Unknown).

Prev cesarian - This field indicates whether previous cesarean delivery is reported as a maternal risk factor (1- Y , 2 – N, 9- Unknown).

Gonorrhea - Gonorrhea, colloquially known as the clap, is a sexually transmitted infection caused by the bacterium Neisseria gonorrhoeae (1- Y , 2 – N, ' ' - Unknown).

Syphilis - Syphilis is a sexually transmitted infection caused by the bacterium Treponema pallidum subspecies pallidum (1- Y , 2 – N, ' ' - Unknown).

Chlamydia - Chlamydia, or more specifically a chlamydia infection, is a sexually transmitted infection caused by the bacterium Chlamydia trachomatis (1- Y , 2 – N, ‘ ‘ - Unknown).

Hepatitis_b – Hepatitis B is a potentially life-threatening liver infection caused by the hepatitis B virus (1- Y , 2 – N, ‘ ‘ - Unknown).

Hepatitis_c - Hepatitis C virus causes both acute and chronic infection. Acute HCV infections are usually asymptomatic and most do not lead to a life-threatening disease (1- Y , 2 – N, ‘ ‘ - Unknown).

Labor_induction - Induction of Labor received by the mother before delivery (1- Y , 2 – N, ‘ ‘ - Unknown).

Labor_augmentation - Augmentation of Labor received by the mother before delivery (1- Y , 2 – N, ‘ ‘ - Unknown).

Steroids - Steroids received by the mother before delivery (1- Y , 2 – N, ‘ ‘ - Unknown).

Antibiotics - Antibiotics received by the mother before delivery (1- Y , 2 – N, ‘ ‘ - Unknown).

Chorioamnionitis – Chorioamnionitis is a serious condition in pregnant women in which the membranes that surround the fetus and the amniotic fluid are infected by bacteria. Yes - clinical chorioamnionitis or maternal temperature >= 38 degrees Celsius (1- Y , 2 – N, ‘ ‘ - Unknown).

Anesthesia - Anesthesia epidural or spinal anesthesia during labor (1- Y , 2 – N, ‘ ‘ - Unknown).

Apgar5 - The "apgar score 5 " is an examination performed on a baby immediately after 5 minutes birth(0-10, ‘ ‘ - Unknown)

Apgar10 - The "apgar score 5 " is an examination performed on a baby immediately after 10 minutes birth(0-10,88- Not Applicable, ‘ ‘ - Unknown)

Plurality - The Plurality field indicates if more than one infant shared the gestation and birth. (1-5 or more).

Gender – Gendr of infant(F – Female , M – Male)

Infant_weight - The Infant Birth Weight field indicates weight ranges for the infant at birth in grams (27-7938,9999- Unknown, ‘ ‘ - Unknown).

Statistics of data

	count	mean	std	min	25%	50%	75%	max
id	738174.0	92005.927632	53929.517272	1.0	45412.0	71123.0	137368.0	197806.0
birth_year	738174.0	2017.932447	1.418983	2016.0	2017.0	2018.0	2019.0	2020.0
birth_month	738174.0	6.525621	3.404681	1.0	4.0	7.0	9.0	12.0
birth_time	738174.0	1230.985486	628.472775	0.0	801.0	1238.0	1732.0	2359.0
birth_place	738174.0	1.029807	0.271642	1.0	1.0	1.0	1.0	7.0
mother_age	738174.0	29.280554	5.645909	13.0	25.0	29.0	33.0	50.0
mother_education	738174.0	4.522692	1.738562	1.0	3.0	4.0	6.0	8.0
father_age	738174.0	31.673272	6.765566	11.0	27.0	31.0	36.0	91.0
father_education	738174.0	4.256521	1.746747	1.0	3.0	4.0	6.0	8.0
interval_llb	738174.0	376.970744	410.483639	3.0	32.0	75.0	888.0	888.0
cigarettes	738174.0	0.893218	4.253891	0.0	0.0	0.0	0.0	98.0
mother_height	738174.0	64.150585	2.828501	39.0	62.0	64.0	66.0	78.0
mother_bmi	738174.0	27.052399	6.645042	13.1	22.2	25.6	30.5	69.3
pre_preg_weight	738174.0	158.533460	41.151141	75.0	130.0	150.0	180.0	375.0
delivery_weight	738174.0	188.073227	40.521529	100.0	160.0	181.0	210.0	400.0
prev_cesarian	738174.0	0.001733	0.124820	0.0	0.0	0.0	0.0	9.0
apgar5	738174.0	8.801239	0.758388	0.0	9.0	9.0	9.0	10.0
apgar10	738174.0	87.153468	8.284783	0.0	88.0	88.0	88.0	88.0
plurality	738174.0	1.032227	0.182279	1.0	1.0	1.0	1.0	5.0
infant_weight	738174.0	3290.905058	583.604595	227.0	2990.0	3320.0	3648.0	9999.0

Unique values each

id	196553
birth_year	5
birth_month	12
birth_time	1440
birth_place	7
mother_age	38
marital_status	3
mother_education	8
father_age	71
father_education	8
interval_llb	298
cigarettes	54
mother_height	38
mother_bmi	547
pre_preg_weight	301
delivery_weight	301
pre_preg_diabetes	2
gest_diabetes	2
pre_preg_hypertension	2
gest_hypertension	2
prev_preterm_birth	2
infertility_treatment	2
prev_cesarian	3
gonorrhea	2
syphilis	2
chlamydia	2
hepatitis_b	2
hepatitis_c	2
labor_induction	2
labor_augmentation	2
steroids	2
antibiotics	2
chorioamnionitis	2
anesthesia	2
apgar5	11
apgar10	12
plurality	5
gender	2
infant_weight	4518

Variables

Variables continuous – [birth_time, mother_age, interval_llb, father_age, mother_height, mother_bmi, pre_preg_weight, delivery_weight, infant_weight]

Variables categorical- [birth_year, birth_month, birth_place, marital_status, mother_education, father_education]

Variables binary- [pre_preg_diabetes, gest_diabetes, pre_preg_hypertension, gest_hypertension, prev_preterm_birth, infertility_treatment, prev_cesarian, gonorrhea, syphilis, chlamydia, hepatitis_b, hepatitis_c, labor_induction, labor_augmentation, steroids, antibiotics, chorioamnionitis, anesthesia, gender]

Dtype each columns

id	Int64
birth_year	Int64
birth_month	Int64
birth_time	Int64
birth_place	Int64
mother_age	Int64
marital_status	string
mother_education	Int64
father_age	Int64
father_education	Int64
interval_llb	Int64
cigarettes	Int64
mother_height	Int64
mother_bmi	Float64
pre_preg_weight	Int64
delivery_weight	Int64
pre_preg_diabetes	string
gest_diabetes	string
pre_preg_hypertension	string
gest_hypertension	string
prev_preterm_birth	string
infertility_treatment	string
prev_cesarian	Int64
gonorrhea	string
syphilis	string
chlamydia	string
hepatitis_b	string
hepatitis_c	string
labor_induction	string
labor_augmentation	string
steroids	string
antibiotics	string
chorioamnionitis	string
anesthesia	string
apgar5	Int64
apgar10	Int64
plurality	Int64
gender	string
infant_weight	Int64

There is number_of_Features : 39

Example of plots

