



# 피마 인디언 당뇨병 진단 프로젝트

2조: 김가현, 김수현, 박인혜, 방채은

30602김가현: 웹 사이트제작  
30607김수현: 아바타제작  
30616박인혜: ppt제작  
30617방채은: orange3마무리



# 목차

1. 기획
2. 데이터
3. 시각화
4. 모델
5. 웹
6. 개선점&기대점
8. 마무리 인사



# 사용 툴 모음

오렌지 3  
클로버 AI  
VS코드



# 0. 기획 이유★





## 0. 기획 이유

기획 이유:  
조원들의 진로  
분야와 관련되어  
보이는 '파마산  
인디언 당뇨병'을  
주제로 선정



# 1. 사용한 데이터 설명



## 사용한 데이터

- 1) Pregnancies: 임신 횟수
- 2) Glucose: 2시간 동안의 경구 포도당 내성 검사에서 혈장 포도당 농도
- 3) Blood Pressure: 이완기 혈압(mm Hg)
- 4) Skin Thickness: 삼두근 피부 주름 두께 (mm)
- 5) Insulin: 2시간 혈청 인슐린 수치(mu U/ml)
- 6) BMI: 체질량 지수 ( $BMI = \text{체중(kg)} / \text{키(m)}^2$ )
- 7) Diabetes Pedigree Function: 당뇨병 혈통 기능
- 8) Age: 나이(세)

9) Outcome: 결과 (당뇨병이 아니면 0,



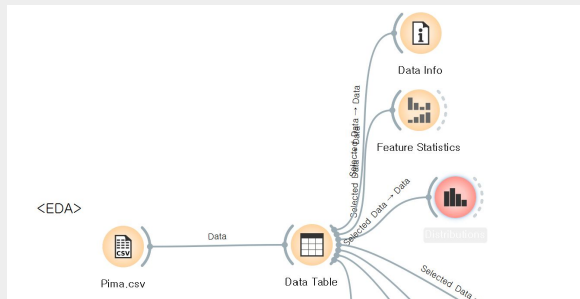


## 2. 데이터 탐색

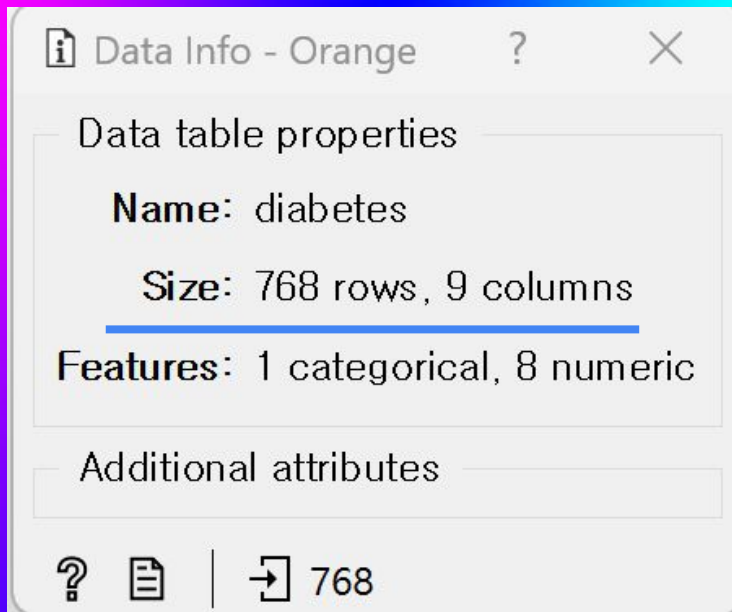


# 데이터 탐색

Data info, Feature Statistics,  
Distributions, Scatter Plot, Box  
Plot, Violin Plot을 사용하여 데이터  
시각화



# Data info-행&열 탐색



- 768개의 가로 줄 (행, instances)
- 9개의 세로 줄 (열, features)

# Data table-데이터 전체 살펴봄

Data Table - Orange

Info  
768 instances (no missing data)  
9 features  
No target variable.  
No meta attributes.

Variables  
☒ Show variable labels (if present)  
☐ Visualize numeric values  
☒ Color by instance classes

Selection  
☒ Select full rows

Restore Original Order

☒ Send Automatically

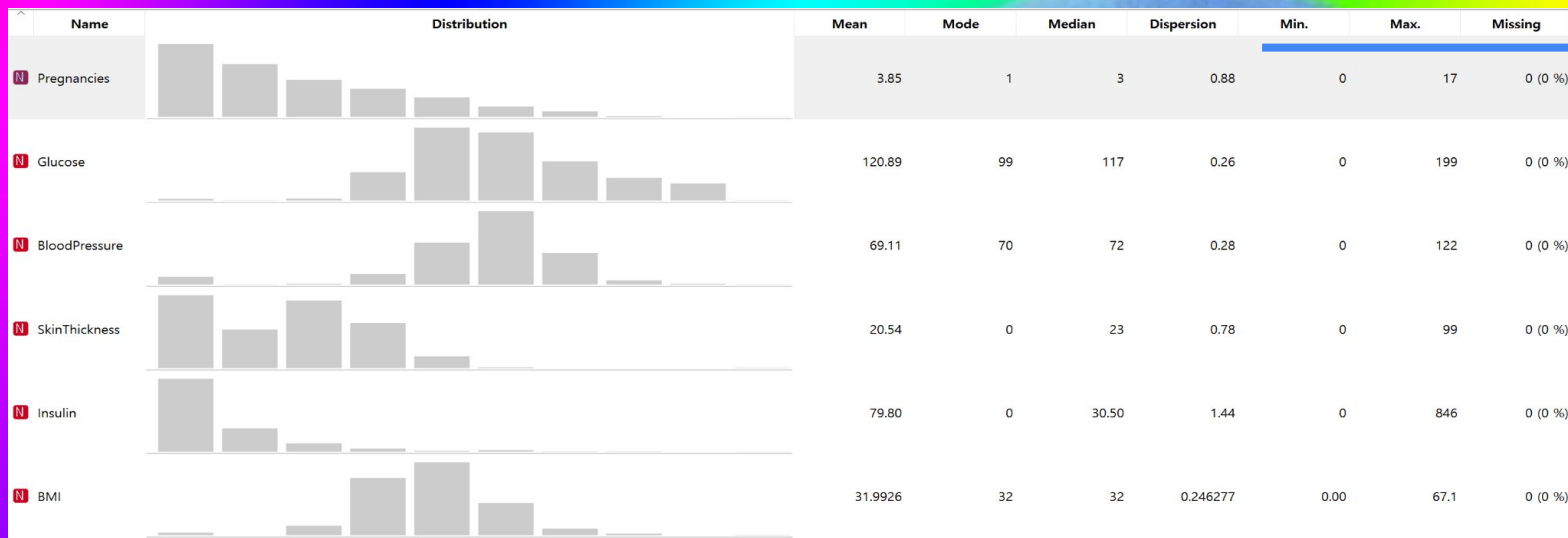
	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	atesPedigreeFun	Age	Outcome
748	1	81	74	41	57	46.3	1.096	32	0
749	3	187	70	22	200	36.4	0.408	36	1
750	6	162	62	0	0	24.3	0.178	50	1
751	4	136	70	0	0	31.2	1.182	22	1
752	1	121	78	39	74	39	0.261	28	0
753	3	108	62	24	0	26	0.223	25	0
754	0	181	88	44	510	43.3	0.222	26	1
755	8	154	78	32	0	32.4	0.443	45	1
756	1	128	88	39	110	36.5	1.057	37	1
757	7	137	90	41	0	32	0.391	39	0
758	0	123	72	0	0	36.3	0.258	52	1
759	1	106	76	0	0	37.5	0.197	26	0
760	6	190	92	0	0	35.5	0.278	66	1
761	2	88	58	26	16	28.4	0.766	22	0
762	9	170	74	31	0	44	0.403	43	1
763	9	89	62	0	0	22.5	0.142	33	0
764	10	101	76	48	180	32.9	0.171	63	0
765	2	122	70	27	0	36.8	0.34	27	0
766	5	121	72	23	112	26.2	0.245	30	0
767	1	126	60	0	0	30.1	0.349	47	1
768	1	93	70	31	0	30.4	0.315	23	0

768 | 768 | 768

768개의 가로 줄 (행, instances)  
9개의 세로 줄 (열, 칼럼, features)



# Feature Statistics-결측치 확인



- min, max값 확인 할 수 있다.
- missing(결측치) 확인 할 수 있다.



# Scatter Plot-산점도



Scatter Plot

## Scatter Plot(산점도)

- 두 변수 사이의 상관 관계를 나타내는 그래프
- 산점도는 주로 연속형 변수를 사용하여 데이터 포인트를 표시함.  
x축은 한 변수의 값을, y축은 다른 변수의 값을 나타낸다.  
이렇게 그래프를 그리면 데이터 간의 패턴이나 관계를 시각적으로 파악할 수 있다.
- 예시) x축에는 학생들의 공부 시간을, y축에는 그들의 시험 점수를 나타내는 데이터가 있다고 가정.  
각 학생을 하나의 점으로 표현하여 공부 시간과 시험 점수 간의 관계를 살펴볼 수 있다.  
만약 데이터 포인트들이 대체적으로 오른쪽 위쪽으로 향하고 있다면,  
공부 시간이 증가할수록 시험 점수도 증가한다는 양적 관계가 있을 수 있다.  
반대로, 데이터 포인트들이 대체적으로 왼쪽 아래쪽으로 향하고 있다면,  
공부 시간이 증가할수록 시험 점수는 감소한다는 음적 관계가 있을 수 있다.
- 산점도는 이러한 상관 관계뿐만 아니라 이상치나 군집(clustering) 등의 패턴을 파악하는 데에도 유용하다.  
데이터 분포를 시각화하여 데이터 간의 관계를 직관적으로 이해할 수 있도록 도와준다.

# Feature Statistics-결측치 확인



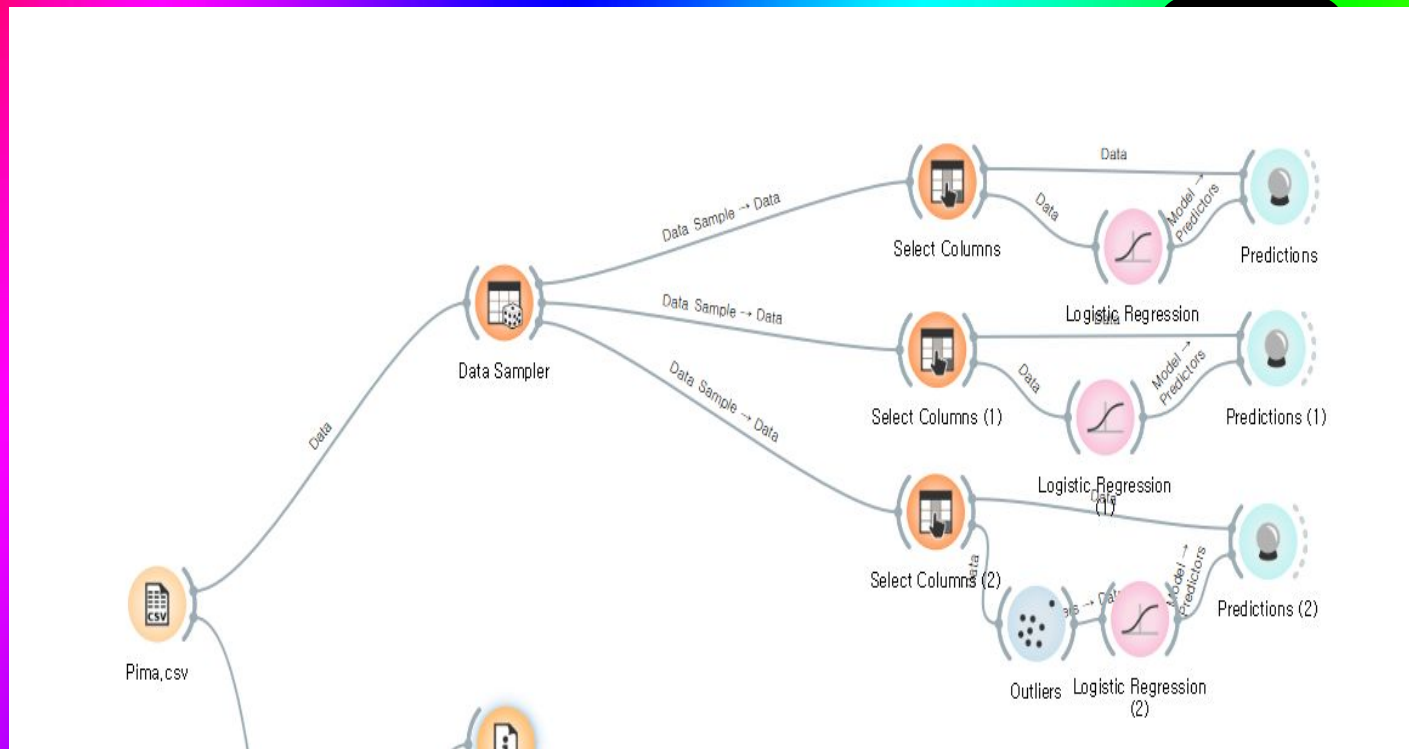
Name	Distribution	Mean	Mode	Median	Dispersion	Min.	Max.	Missing
Pregnancies		3.85	1	3	0.88	0	17	0 (0 %)
Glucose		120.89	99	117	0.26	0	199	0 (0 %)
BloodPressure		69.11	70	72	0.28	0	122	0 (0 %)
SkinThickness		20.54	0	23	0.78	0	99	0 (0 %)
Insulin		79.80	0	30.50	1.44	0	846	0 (0 %)
BMI		31.9926	32	32	0.246277	0.00	67.1	0 (0 %)



## 4. 데이터 모델링



# 4. 데이터 모델링



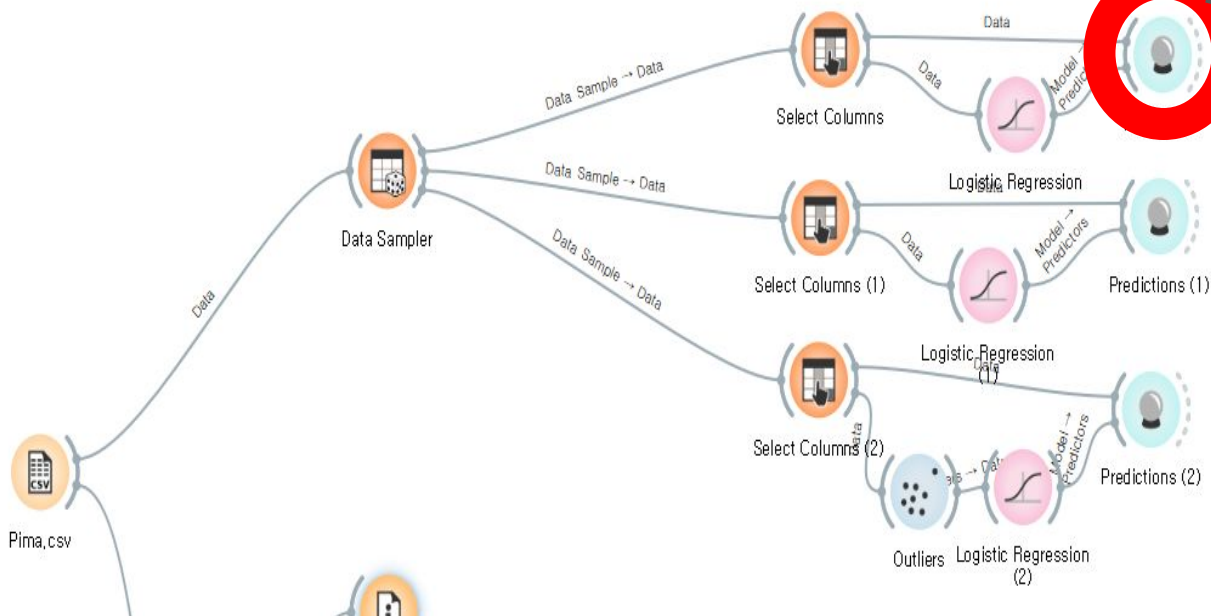
# 5. 결과 발표





# 5. 결과 발표

가장  
정확도가  
높은 모델로  
선택



## 6. 향후 개선점, 기대점



## 향후 개선점, 기대점

인디언 당뇨병에 영향을 주는  
요인들을 알 수 있었고 또 다른  
요인으로 상관관계를 오렌지3를  
이용하여 모델링해볼 수 있을 것  
같다.



감사합니다



“오렌지3를 이용해서 다양한 데이터를 다양한  
표와 그래프로 알아볼 수 있었다.”

