Program ‘Analiza\_danych.py’ to aplikacja umożliwiająca badanie i wizualizację danych statystycznych zawartych w bazie danych ‘dataset\_zadanie.csv’

Po uruchomieniu aplikacji użytkownik ma dostęp do kilku funkcjonalności, które można wybrać poprzez wpisanie właściwego indeksu i zatwierdzenie enterem.

**Model predykcyjny**

Model dokonuje analizy bazy danych i przewiduje zmienną outcome.

Program przewiduję 300 zmiennych Outcome dla 300 pierwszych przypadków i zapisuje je w pliku ‘outcome\_prediction.txt’. Przewidziane zmienne zostały przeniesione do pliku ‘dataset\_zadanie.xlsx’

Model predykcyjny oblicza zmienną Outcome na podstawie 'Age', 'Glucose' i 'Insulin', ponieważ te zmienne mają największy wpływ na zmianę Outcome – obliczone używając rozkładu chi^2 w funkcji o nazwie ‘predict\_outcome’ znajdującej się w kodzie.

Do utworzenia programu użyte zostały następujące biblioteki:

pandas

numpy

matplotlib

scikit-learn