# PAD 09

**Praca Domowa 9 – Testy statystyczne i modelowanie statystyczne**

1. Przed rozpoczęciem zainstaluj biblioteki scipy i statmodels poleceniem:

*pip install NAZWA*  lub *conda install NAZWA*

**Zadanie 1 (3 pk)**

Przeprowadź test T studenta na danych w pliku PAD\_PD\_09.csv

Hipoteza zerowa brzmi: nie ma istotnej różnicy w średnich Annual Income pomiędzy kobietami i mężczyznami.

Hipoteza alternatywna brzmi: istnieje istotna różnica (bez wskazania kierunku) w średnich Annual Income pomiędzy kobietami i mężczyznami.

Przyjmij significance level (alpha) na poziomie 0.05. Czy uda się odrzucić hipotezę zerową?

Pomocne wideo: <https://www.khanacademy.org/math/ap-statistics/xfb5d8e68:inference-quantitative-means/two-sample-t-test-means/v/two-sample-t-test-for-difference-of-means> - przeprowadzanie testu dla dwóch zmiennych.

**Zadanie 2 (4 pkt)**

Korzystając z tych samych danych i biblioteki statmodels stwórz model regresji liniowej, gdzie zmienną zależną jest spending score, a zmiennymi niezależnymi pozostałe zmienne.

Z modelu wyświetl p values, standard dev i coefficients.

Sprawdź czy istnieje korelacja pomiędzy zmiennymi. Pokaż to na wykresie, np. z biblioteki plotly express.

Korzystając z eliminacji wstecznej usuń najmniej istotną zmienną i jeszcze raz wyświetl model.

**UWAGA: zwróć uwagę na zmienne będące kategoriami**!