Marcin Lew 289052

Kajetan Dreliszak 289050

Raport projekt 2 TASS

**Temat 4:** Wyszukiwarka popularnych miejsc turystycznych: na podstawie danych o trasach lotów i danych demograficznych znaleźć miasta o nieproporcjonalnie dużym w stosunku do liczby mieszkańców ruchu lotniczym.

# Fuzja danych

Na podstawie połączonych danych wyznaczono współczynnik liczony jako stosunek liczby pasażerów obsłużonych przez lotniska w okolicach miasta do liczby mieszkańców danego miasta. W efekcie otrzymano prawie 649 wartości dla różnych miast z całego świata.

# Wizualizacja

Każde z rozpatrywanych miast zostało zaznaczone na mapie *Google’a* za pomocą koło w konkretnym kolorze oraz etykiety przypiętej do współrzędnych geograficznych miasta. Wartość uwzględniono w trojaki sposób:

* Koła przyjmuje jedną z następujących barw:
  + fiolet,
  + niebieski
  + zielony,
  + żółty,
  + pomarańczowy,
  + czerwony.

Gdzie kolor fioletowy przypada miastom o małych współczynnikach , a czerwony tym o dużych. Przy czym za wartość graniczną , po której barwa nie ulega zmianie, przyjęto .

* Wraz ze wzrostem rośnie promień koła. Podobnie jak w przypadku kolorów, miasta o wartości większej niż mają koła jednakowej wielkości.
* Etykieta posiada taki sam kolor, jak koło. Ponadto najeżdżając na nią wyświetla się nazwa miasta oraz dokładna wartość .

# Analiza wartości wskaźnika

Podczas analizy brano pod uwagę miasta z całego świata, jednak na potrzeby tego raportu zawężono opisywany obszar do Europy.  
Z początku dane naniesione na mapę wydają się spełniać założenia projektu – kolor czerwony i pomarańczowy dominują: w okolicach basenu morza Śródziemnomorskiego, w kilku skandynawskich miastach, na wyspach Kanaryjskich oraz na innych wyspach takich jak Sardynia, Kostaryka i Malta czy też w większych europejskich stolicach jak Paryż czy Amsterdam.   
Zauważono jednak, że poza miejscami, które znane są dużej liczby turystów, współczynnik wyróżnił również mniej turystyczne miejscowości, takie jak Billund (Dania, ), Shannon (Irlandia,), Faro (Portugalia, ), Keflavik (Islandia, ) czy Frankfurt nad Menem (Niemcy, ). Odpowiedź na pytanie, dlaczego tak się stało, leży w specyficznych cechach tych miast, których nie uwzględnia współczynnik :

* mała liczba ludności, stosunkowo duże lotnisko (Billund, Faro, Shannon). Billund, z powodu *Legolandu*, jest miejscowością turystyczną, ale zdecydowanie mniej niż wskazuje na to .
* lotnisko przesiadkowe (tranzytowe), będące np. łącznikiem pomiędzy Europą a USA) – rzadko kiedy bywa portem docelowym np. Faro.
* przypadek miasta Keflavik – znajduje się tam główny międzynarodowy port lotniczy w Islandii, co w połączeniu z populacją miasta nie przekraczającą 14tyś. Daje bardzo duży współczynnik .

# Próba udoskonalenia współczynnika

W celu usunięcia niedoskonałości wynikających z budowy współczynnika zdecydowano się uwzględnić kolejny rodzaj danych – informacje o istniejących połączeniach pomiędzy lotniskami. Na ich podstawie stworzono graf skierowany *DiGraph* i przystąpiono do jego analizy.

Początkowo policzono współczynnik gronowania (ang. *clustering coefficient*) dla każdego z portów – węzły grafu. Następnie przemnożono przez niego i wynik naniesiono na mapę. Operacja ta przyniosła jednak dwojakie efekty. Znacznie zmalał dla miasta Frankfurt nad Menem. Jednak dla Faro, Shannon i Billund dysproporcja w stosunku do innych miast znacznie wzrosła. Ponadto na mapie znacznie „urosło” miasto Kirkenes, które znajduje się w północnej części Norwegii i zdecydowanie nie należy do miast turystycznych. Dlatego też zrezygnowano ze współczynnika gronowania.

Kolejnym krokiem było policzenie pośrednictwa(ang. *closeness*). Pomysł był następujący – miasta pełniące rolę mostów pomiędzy grupami miast zostaną uwydatnione a następnie usunięte. W zamyśle miały być to porty tranzytowe. Okazało się jednak, że współczynnik pośrednictwa nie tylko portów przesiadkowych, a porty centralne w poszczególnych krajach – Paryż, Amsterdam, Madryt.

Próbowano jeszcze wykluczyć z sieci wszystkie przeguby (ang. *articulation\_points*) oraz policzono kilka innych współczynników. Efekty były jednak bardzo niezadowalające i nie zostały w tym raporcie opisane.

# Uruchamianie aplikacji

Aplikację napisano w języku programowania *Python*. Wynik jej działania – mapę z danymi – zapisano w pliku *mapa.html*, który znajduje się w folderze projektu *TASS2*. Plik należy otworzyć w dowolnej przeglądarce (testowane dla *Chrome*).  
Drugą opcją jest uruchomienie skryptu w środowisku *PyCharm*. W wyniku utworzony zostanie nowy plik *mapa.html*, który zostanie automatycznie otworzony w przeglądarce domyślnej użytkownika.

W celu obejrzenia map z uwzględnieniem współczynników omówionych w rozdziale 4 należy otworzyć pliki *mapa\_clustering\_coefficient.html* lub *mapa\_betweenness\_centrality.html* znajdujące się w folderze projektu *TASS2.*

Do wyświetlenia mapy niezbędny jest dostęp do Internetu.