

Programowanie w R – projekt grupowy

e-rejwer@prz.edu.pl

W ramach zajęć przewidziana jest realizacja projektów grupowych. Państwa zadaniem jest wykonanie interesującego opracowania wybranych danych, pochodzących z repozytoriów open data, np.:

European Union Open Data Portal

<https://data.europa.eu/data/datasets?locale=pl&minScoring=0>

National Climatic Data Center

<https://www.ncdc.noaa.gov/data-access/quick-links#loc-clim>

WHO database

<https://www.who.int/data/gho/>

UNICEF Dataset

<https://data.unicef.org/>

World Bank Open Data

<https://data.worldbank.org/>

Kaggle Open Datasets

<https://www.kaggle.com/datasets>

freeCodeCamp Open Data

<https://github.com/freeCodeCamp/open-data>

Projekty grupowe powinny być realizowane w zespołach 2-osobowych, bądź, po mojej akceptacji, w zespołach 3-osobowych. W zespoły dobierają się Państwo samodzielnie. Członkowie zespołu mogą należeć do różnych grup laboratoryjnych/projektowych, o ile wszyscy członkowie grupy będą mogli być obecni podczas referowania projektu.

Starosta grupy lub ochotnik, któremu starosta powierzy to zadanie, jest zobowiązany stworzyć listę tematów projektów, w celu uniknięcia takiej samej lub bardzo zbliżonej tematyki. (Najlepiej, aby Państwo udostępnili sobie listę tematów i informowali siebie na bieżąco o jej modyfikacjach). Będzie obowiązywała jedna lista dla wszystkich grup projektowych. Decydować powinna kolejność zgłaszanych tematów. Na liście tematów powinny znaleźć się 3 elementy: skład zespołu, krótka nazwa lub opis planowanego projektu oraz adres do strony, z której pobrać można dane źródłowe:

Imiona i nazwiska członków zespołu projektowego

"Tytuł projektu" lub krótki opis projektu

Adres strony internetowej, z której pobrano dane.

Kompletną listę proponowanych tematów należy przesłać do mnie do 22 maja 2024r.

Tytuł projektu może być bardzo ogólny i nie musi precyzować aspektów planowanej analizy i opracowania. Temat przedstawiony w końcowym opisie projektu może zostać przeformułowany i doprecyzowany. Podczas oceny projektów znaczenie będzie miała złożoność zrealizowanej analizy, prawidłowość oraz jakość realizacji poniższych elementów oraz brak wskazań sugerujących pracę niesamodzielną.

Realizacja projektu i jego zawartość powinna obejmować i uwzględniać następujące elementy:

Część 1 Zestaw danych

Wybrane zestawy danych powinny posiadać łącznie nie mniej niż 30 000 rekordów (wierszy). Łączna ilość zmiennych (kolumn) nie powinna być mniejsza niż 6. Przy czym zmienne powinny być reprezentowane przez różne typy danych: wektory liczbowe (numeric), znakowe (character), czynniki (factor).

Część 2 Wczytanie danych

Dane najlepiej wczytać do R z oryginalnych plików, dopuszcza się jednak ich wstępną *techniczną obróbkę* w innych programach (np. Excel).

Część 3 Obróbka danych

Tworzony kod powinien opierać się na większości typów danych poznanych na zajęciach, w tym nie mniej niż na trzech ramkach danych, grupujących istotne, z punktu widzenia tematyki projektu, dane. W ramach przygotowania i obróbki danych – agregowania, filtrowania, sortowania, itd. należy wykorzystywać metody poznane na zajęciach. Jest to kluczowy element w całym projekcie, dlatego im więcej zostanie użytych technik (poleceń), tym lepiej. (Policzenie wyłącznie średniej wartości z rekordów spełniających jakieś kryterium, to zbyt mało). W poszukiwaniu pomysłów na realizację projektu warto przejrzeć tematykę podejmowaną na wykładach i zajęciach projektowych.

Część 4 Wizualizacja danych

Podczas realizacji projektu należy skorzystać z technik wizualizacji danych w R. Projekty będą wówczas z pewnością atrakcyjniejsze.

Część 5 Pliki wyjściowe

Projekt grupowy (opisy + kod) proszę przygotować w R Markdown.

W tworzonym kodzie powinny być generowane wyniki prezentujące efekty analizy przeprowadzonej na poszczególnych etapach projektu. W tym celu powinny być "drukowane" jedynie fragmenty danych wyjściowych. (Inaczej wyniki będą mało czytelne). Natomiast istotną rolę powinny stanowić wnioski (podsumowania) z przeprowadzonej analizy, które powinny być jasne i czytelne.

Część 6 Kod z komentarzami

W obrębie kodu należy dołączać komentarze opisujące i numerujące istotne etapy realizacji projektu. (Znowu, im więcej, tym lepiej, ale bez przesady ;)).

Część 7 Opis projektu, wnioski i dziennik z realizacji projektu

Ważną częścią projektu jest opis sposobu realizacji etapów projektu - **Dziennik z realizacji projektu**. **W oddzielnym pliku .doc należy na bieżąco, chronologicznie, krótko i treściwie opisywać kolejne etapy realizacji projektu, ze szczególnym uwzględnieniem napotkanych trudności, nieudanych pomysłów na ich rozwiązanie i sposobów ich rozwiązania lub ominięcia.** Jeśli jakiegoś pomysłu nie udało się zrealizować lub rozwiązać problemu, informacja o nim powinna znaleźć się w dzienniku. Ilość stron dziennika będzie zależała od rozważanego problemu, nie powinna jednak przekroczyć 15 stron.

Część 8 Prezentacja projektu

W ramach prezentacji projektu powinien zostać przygotowany *pokazowy* dokument, dopracowany typograficznie i graficznie, przystosowany do wydrukowania jako plakat w formacie A3, zawierający krótki opis realizowanego zadania oraz prezentujący wnioski z niego wynikające. Dokument powinien podkreślać szczególnie te aspekty, które decydują o atrakcyjności projektu oraz pokazywać najważniejsze (najciekawsze) wyniki jakie osiągnęli Państwo podczas przeprowadzonej analizy. Będzie to bardzo ważny element prezentacji projektu !

Przygotowanego plakatu nie należy drukować !

Termin złożenia projektów: 9 czerwca 2024r.