

R+Excel w analizie danych

Grupa wykładowa

informatyka, studia niestacjonarne, studia magisterskie

semestr 3, rok akademicki 2022 / 2023, Uniwersytet Ekonomiczny w Katowicach

Skład grupy projektowej

- Paweł Piątek
- Karol Pilot
- Tomasz Janus
- Artur Leśnik

Cel projektu

Celem projektu jest poznanie narzędzi skryptowych, które umożliwiają automatyczne pobieranie danych z różnych źródeł, przetwarzają te dane, a następnie generują plik Microsoft Excel, który zawiera przetworzone dane oraz ich reprezentację w formie graficznej. Ma to na celu nauczanie się, jak tworzyć w sposób automatyczny powtarzalne raporty.

Temat projektu

Stworzenie aplikacji pogodowej generującej raport temperatur dla miasta Katowice na nadchodzące 7 dni kalendarzowych.

Opis realizacji projektu

Po zawiązaniu grupy projektowej dokonaliśmy wyboru źródła danych, które będzie dostarczało informacji pogodowych do naszej aplikacji. Wybór padł na API dostarczane przez OpenWeatherMap, ponieważ ta usługa oferuje darmowy pakiet, który pozwala na pobieranie danych pogodowych dla miasta Katowice na nadchodzące 8 dni w formacie JSON.

Kolejnym krokiem był wybór języka programowania. Ponieważ nikt z naszej grupy nie miał wcześniej styczności z językiem R, postanowiliśmy skorzystać z opcji napisania aplikacji w języku Python. Do tego celu wykorzystaliśmy dwie biblioteki Requests oraz OpenPyXL. Pierwsza biblioteka umożliwiła nam przesyłanie zapytania do serwera API z prośbą o dane pogodowe, a druga biblioteka służy do generowania i modyfikowania plików Microsoft Excel przy pomocy języka Python.

Pisanie aplikacji rozpoczęliśmy od napisania klasy pobierającej dane z serwera API. Zapytanie kierowane jest do serwera API, jako parametry podawane są współrzędne geograficzne miasta

Katowice oraz klucz API przypisany do naszego darmowego konta. Dane udało się pobrać pomyślnie, więc stworzyliśmy nową tabelę, w której przechowywaliśmy tylko te kolumny, które będą nam potrzebne do wygenerowania odpowiedniego wykresu.

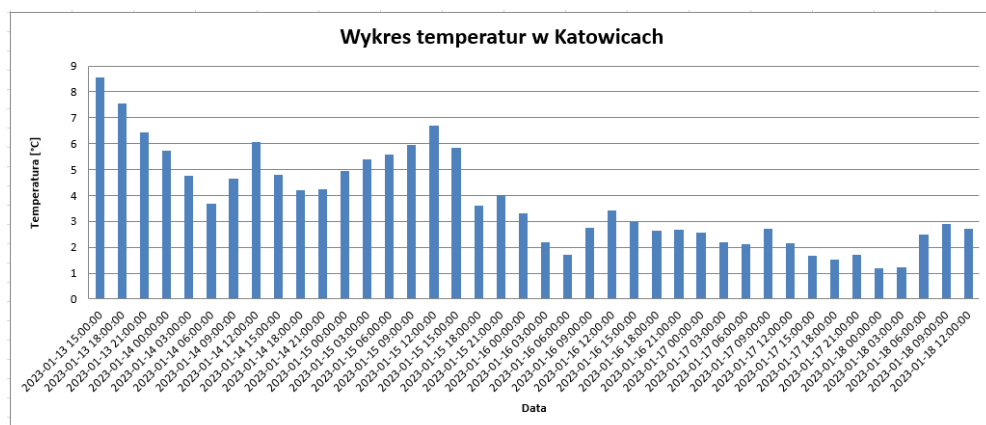
Kolejnym krokiem było stworzenie pustego pliku Excel przy pomocy biblioteki OpenPyXL. Do wykresu dodaliśmy nowy arkusz, a w nim tabelę z danymi do której dodaliśmy nagłówki opisujące znaczenie poszczególnych danych. Zapisana tabela prezentuje się następująco:

| Date | Temperature |
|---------------------|-------------|
| 2023-01-13 15:00:00 | 8,57 |
| 2023-01-13 18:00:00 | 7,56 |
| 2023-01-13 21:00:00 | 6,45 |
| 2023-01-14 00:00:00 | 5,73 |
| 2023-01-14 03:00:00 | 4,75 |
| 2023-01-14 06:00:00 | 3,68 |
| 2023-01-14 09:00:00 | 4,66 |
| 2023-01-14 12:00:00 | 6,06 |
| 2023-01-14 15:00:00 | 4,78 |

W pierwszej kolumnie znajduje się oznaczenie czasu prognozy pogody, a w drugiej kolumnie znajduje się prognozowana temperatura w stopniach Celsjusza. Kolejnym krokiem było poddanie tych danych obróbce statystycznej. Postanowiliśmy, że dla całej tabeli wyliczmy średnią oraz wartości minimalne i maksymalne. W tym celu programowo dodaliśmy trzy nowe wiersze, które zawierają formuły programu Excel do wyliczenia tych parametrów. Wyliczone dane prezentują się następująco:

| | |
|---------------------|------|
| Average temperature | 3,79 |
| Max temperature | 8,57 |
| Min temperature | 1,18 |

Ostatnim krokiem było programowe dodanie wykresu do arkusza, zdecydowaliśmy, że dodamy go do tego samego arkusza zamiast tworzyć nowy specjalnie dla niego. Nasz wykres prezentuje się następująco:



Ostatnim krokiem wykonanym programowo jest zapisanie wyniku naszej pracy do pliku XLSX. Każda osoba, która uruchomi nasz program jest w stanie wygenerować sobie taki plik bez żadnej pomocy. Tym samym osiągnęliśmy wszystkie zakładane przez nas cele dla tego projektu.

Wnioski końcowe z projektu

Poprzez realizację naszego projektu, uznajemy, że osiągnęliśmy cel ćwiczenia, którym było nauczenie się automatycznego tworzenia raportów ze źródeł zewnętrznych w programie Excel przy pomocy języków programowania. Umiejętności, które pozyskaliśmy pozwolą nam w przyszłości na tworzenie automatycznych powtarzalnych raportów na potrzeby innych osób.

Załączniki

- Do niniejszego raportu został dołączony kod źródłowy stworzonej aplikacji oraz przykładowy raport wygenerowany przez nią.