

AUTOMATED WEB SCRAPING

Introduction | Website Exploration | Data Scraping | Data Analysis | Summary



OUTLINE

Introduction



Website
Exploration



Data Scraping



Data Analysis



Summary



Introduction



About the project:

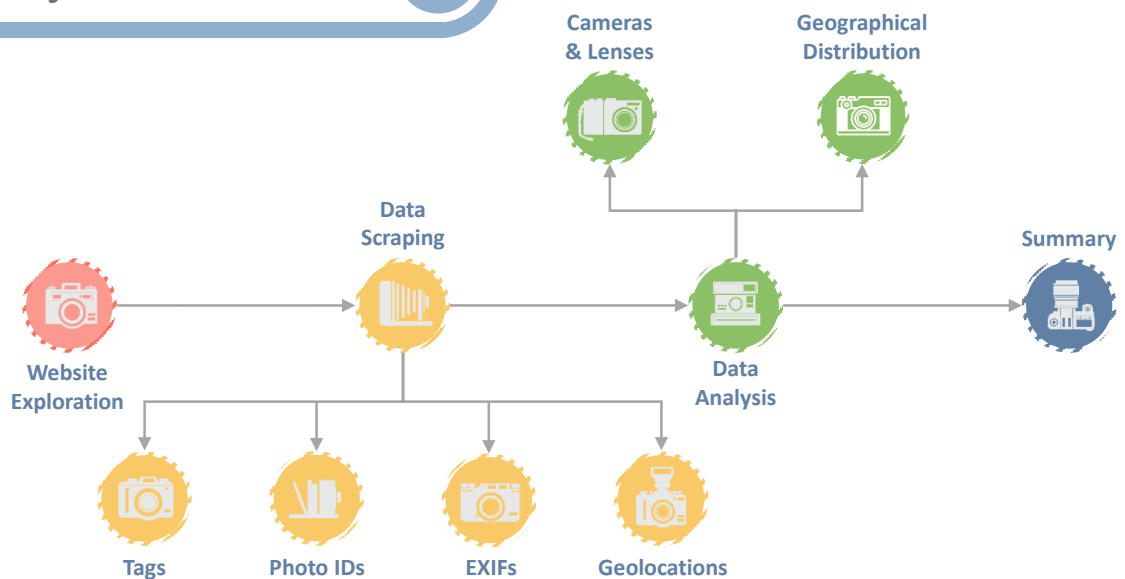
- To build a web scraper for Flickr website.
- To collect photographs and the key elements behind such as:
 - Topics
 - Equipment used
 - Camera settings
 - Geographic locations
- To analyze and visualize the data using Tableau.



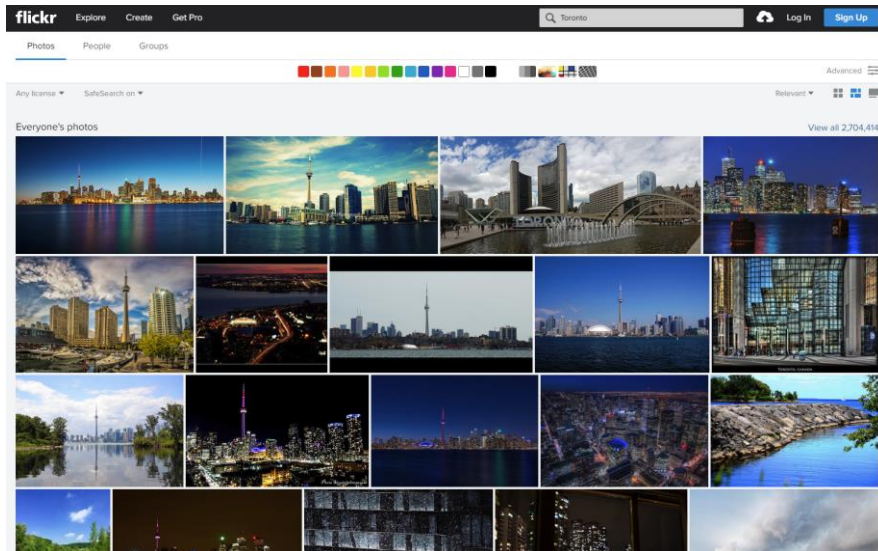
Objective: To give the intuition of which kinds of equipment are applicable for given topics.



Project Workflow



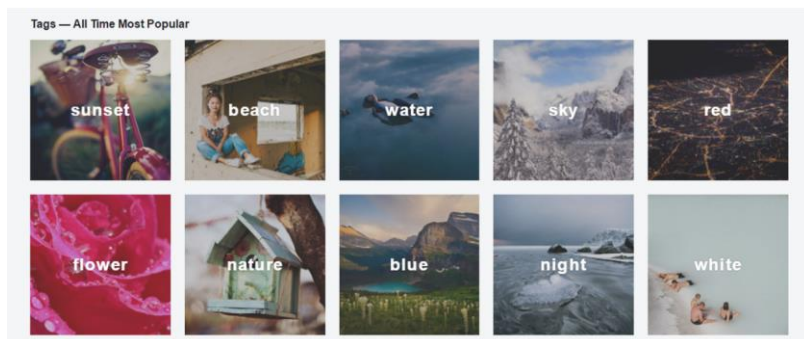
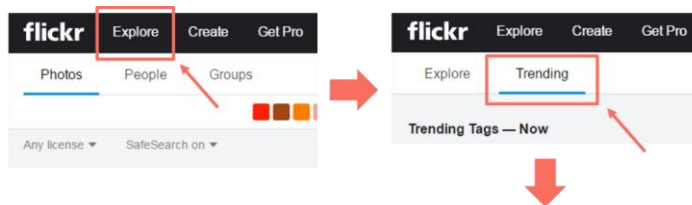
Website Exploration



Website Exploration



The hottest tags:



Website Exploration



Picture details (EXIFs):

The screenshot shows a Flickr photo page for a photo taken on May 24, 2016. The photo is by VV Nincic, a user from Toronto. The photo has 22,064 views, 110 faves, and 12 comments. The photo is titled "Taken on May 24, 2016" and has a "Some rights reserved" license. The photo is a landscape shot of a city skyline. The photo is annotated with red boxes and arrows pointing to various details:

- Views:** 22,064 views, 110 faves, 12 comments
- Equipment:** Nikon D5300, 18-140mm f/3.5-5.6
- Camera settings:** f/6.3, 1/20, ISO 400, Flash (off, did not fire)
- Geolocation:** Toronto, Ontario, Canada
- Detailed settings:** Show EXIF

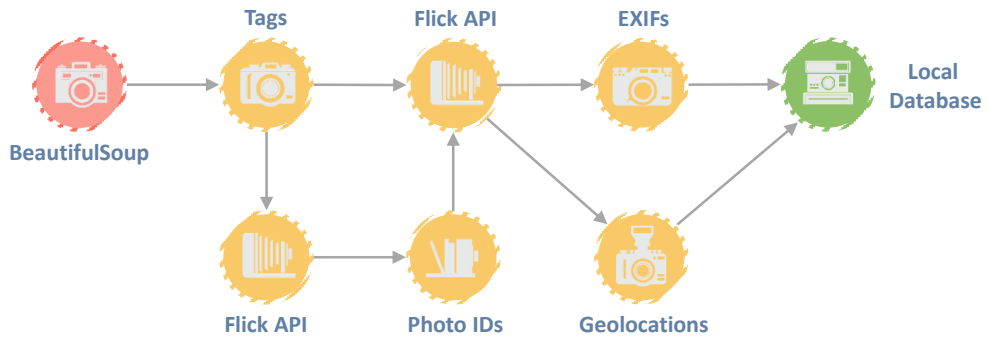
The photo is also annotated with a red box around the "Show EXIF" button, which is labeled "Detailed settings".



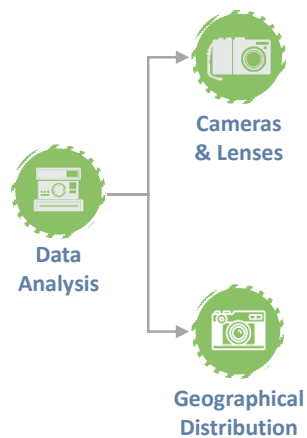
Data Scraping



Scraping Process



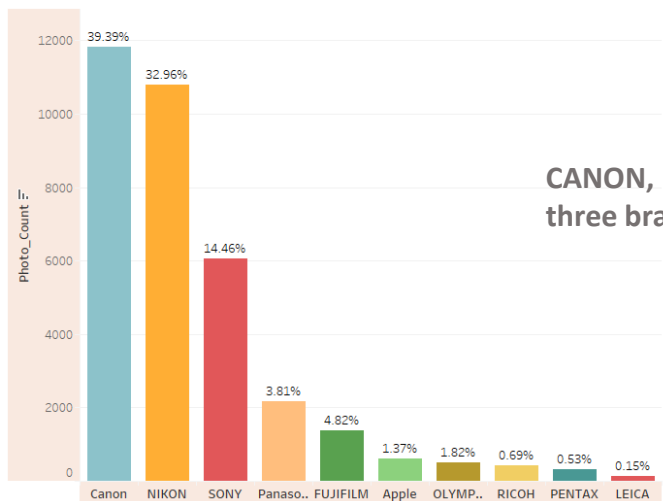
Data Analysis



Data Analysis



Brand ranking



CANON, NIKON and SONY are the top three brands used by Flickr users.

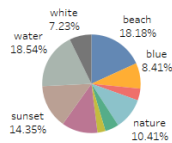


Data Analysis

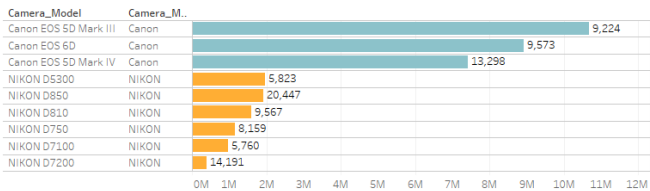


Cameras & Lenses:

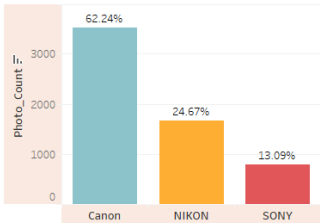
Tags



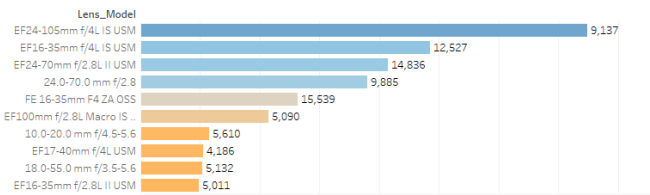
Cameras



Brands



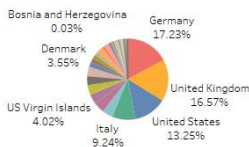
Lenses



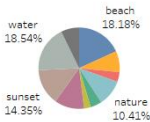
Data Analysis



Geographical distribution:



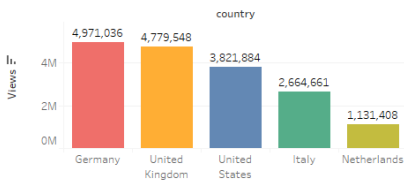
Country - Tag



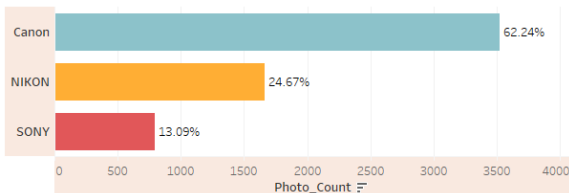
Summary



- Western Countries seem like good places to take good pictures.



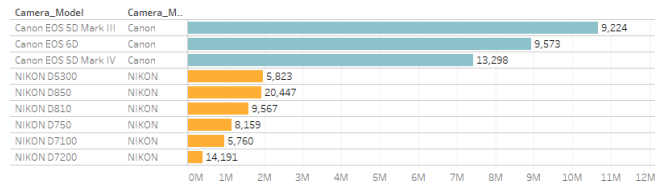
- Top three brands on Flickr are CANON, NIKON, and SONY.



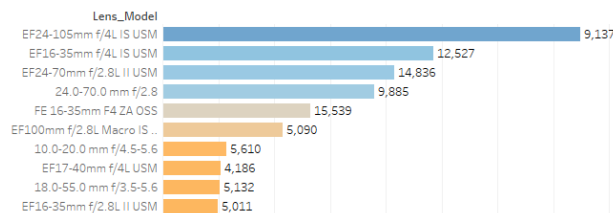
Summary



- Top three camera models are all CANON's products which are a lot more popular than NIKON's cameras.



- The most commonly used lenses are mostly from CANON.



Thank you

