Project Proposal
Introduction to Python

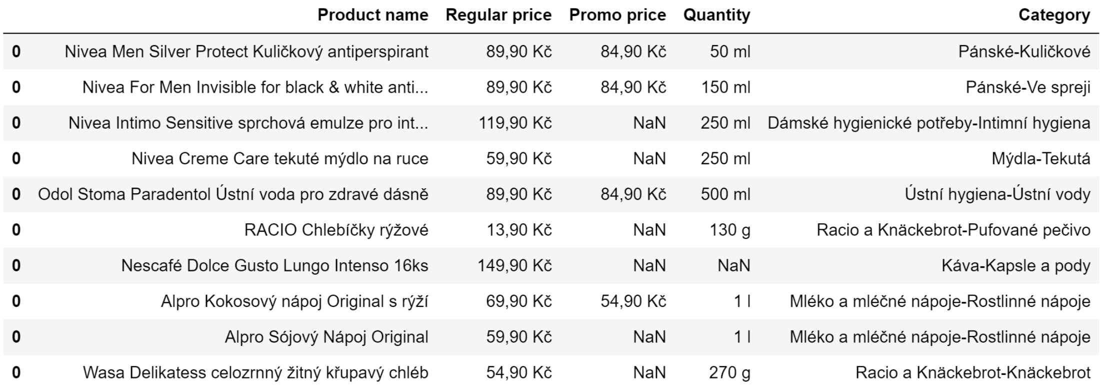
Petra Aschermannová, Aneta Pinteková
51442559@fsv.cuni.cz, 75574795@fsv.cuni.cz
Charles University

March 28, 2019

The goal of this project is to identify products sold by the online shop rohlik.cz and scrape their prices. We would like to accomplish this using the Python packages *Pandas*, *requests*, *xml.etree.ElementTree*.

In the Rohlik.cz webiste, we can find the list of products that rohlik.cz sells. We would like to scrape the prices of these products once a week and then observe (for some selected products, because there are currently around 14 800 products in total) how their prices evolved over the selected period. We are also going to scrape the promotion prices if these are available, volume/quantity of product sold (if available) and category description. We can then do analysis by category. For example, we can inspect how certain categories are being promoted compared to others.

We are going to scrape this information from the website and then save it into a Pandas data frame. An example for the first 10 products is available below:



	Product name	Regular price	Promo price	Quantity	Category
0	Nivea Men Silver Protect Kuličkový antiperspirant	89,90 Kč	84,90 Kč	50 ml	Pánské-Kuličkové
0	Nivea For Men Invisible for black & white anti...	89,90 Kč	84,90 Kč	150 ml	Pánské-Ve spreji
0	Nivea Intimo Sensitive sprchová emulze pro int...	119,90 Kč	NaN	250 ml	Dámské hygienické potřeby-Intimní hygiena
0	Nivea Creme Care tekuté mýdlo na ruce	59,90 Kč	NaN	250 ml	Mýdla-Tekutá
0	Odol Stoma Paradentol Ústní voda pro zdravé dásně	89,90 Kč	84,90 Kč	500 ml	Ústní hygiena-Ústní vody
0	RACIO Chlebíčky rýžové	13,90 Kč	NaN	130 g	Racio a Knäckebrot-Pufované pečivo
0	Nescafé Dolce Gusto Lungo Intenso 16ks	149,90 Kč	NaN	NaN	Káva-Kapsle a pody
0	Alpro Kokosový nápoj Original s rýží	69,90 Kč	54,90 Kč	1 l	Mléko a mléčné nápoje-Rostlinné nápoje
0	Alpro Sójový Nápoj Original	59,90 Kč	NaN	1 l	Mléko a mléčné nápoje-Rostlinné nápoje
0	Wasa Delikatess celozrnný žitný křupavý chléb	54,90 Kč	NaN	270 g	Racio a Knäckebrot-Knäckebrot

However, we are not sure whether this satisfies the ”reproducibility of results” since we cannot go back in time and scrape the prices from e.g. three weeks ago. Nevertheless, *code reproducibility* should be satisfied, *data reproducibility* is probably not necessary.