**0. Administrative Details**

Title: Market Dynamics of Refurbished iPhones: A Web Scraping and Data Analysis Approach

Author: Nico Swart

Supervisor: Dr. Robin Langerak

Institution: Universiteit van Amsterdam

EC: 18

**1. Context of the Project**

In recent years, the market for refurbished electronic devices, especially smartphones like iPhones, has seen significant growth. In the year 2020 for example in the European region the refurbished smartphone market grew by 10%(Counterpoint, 2021). This surge is driven by a combination of economic incentives and environmental benefits, as refurbished devices offer consumers more affordable technology options while also reducing electronic waste. Despite these advantages, the secondary market for such devices remains highly volatile, with prices fluctuating due to factors like technological advances, new product releases, and changing consumer preferences.

This volatility poses a challenge for sellers who must navigate these fluctuating prices to set competitive yet profitable pricing strategies. Currently, the lack of a standardized mechanism for pricing in this market leads to inefficiencies and economic losses. Furthermore, the rapid pace of change in technology and consumer trends demands dynamic pricing models that can respond in real-time to market conditions. However, developing such models requires a deep understanding of market dynamics, which in turn necessitates robust and up-to-date market data.

The marketplace of choice for this study, Marktplaats, is a popular platform for buying and selling secondhand items, including electronics. The platform has a monthly active user base of over 10 million(RTL,2023). It provides a rich dataset due to its extensive user base and high transaction volume. By focusing on data related to iPhone models, specifically storage capacity and pricing, extracted through web scraping techniques, this research aims to analyze and interpret market trends. This analysis will serve as the foundation for proposing a dynamic pricing model that could help sellers of refurbished smartphones optimize their pricing strategies.

The anticipated approach involves the collection and analysis of multiple data points over a period to determine average market prices and their fluctuations. This method will not only help in understanding the current market landscape but also in predicting future trends. The ultimate goal is to leverage this information to support the creation of a pricing model that adjusts dynamically to the ever-changing market conditions, thereby supporting economic sustainability and competitive pricing in the refurbished electronics market.

This research is positioned at the intersection of economic theory and data science, utilizing advanced analytical techniques to address a pressing commercial and environmental issue. The outcomes are expected to contribute valuable insights to the field of applied economics and provide a practical tool for businesses in the refurbished electronics sector.

**2. Research Goal & Relevance**

2.1 Research GoalThe central aim of this research is to analyze and interpret market trends for refurbished iPhones on the Marktplaats platform by focusing on data related to iPhone models, specifically storage capacity and pricing, extracted through web scraping techniques. This comprehensive analysis will identify patterns and fluctuations in market pricing over time, providing a foundational understanding that could inform the development of a pricing strategy guide. While not creating a fully dynamic pricing model, the goal is to offer sellers actionable insights that help them optimize their pricing strategies based on observed market trends.

2.2 Societal RelevanceThe societal impact of this research is significant. By enabling sellers to price refurbished iPhones more effectively, the project supports economic sustainability within the secondary market. It allows technology access at more affordable prices, broadening the demographic that can afford these devices and promoting technological inclusivity. Environmentally, better pricing models and enhanced market efficiency encourage the reuse of electronic devices, contributing to a reduction in electronic waste. This reuse is increasingly important as the global push towards sustainability intensifies, with electronic waste reduction being a critical component of environmental strategies.

2.3 Scientific Relevance  
This study contributes to the scientific fields of applied economics and market analysis by demonstrating how market data can be leveraged to improve economic decisions in the secondary market for electronics. It explores the intersection of data science and economic theory, particularly how empirical data can guide economic strategies in volatile markets. The research will also refine the methodologies used for web scraping and data analysis in non-standardized markets, addressing common challenges such as data inconsistency and noise management.

**3. Research Question**

This thesis proposes a methodical approach to scrape data on refurbished iPhones from the Marktplaats platform and delves into the process required to cleanse this data for analysis. It is designed to establish a robust dataset that serves as a foundation for examining real-time market trends in the refurbished iPhone market.

The study is structured around one main research question and a supporting sub-question. The main question is: "Can a dataset, systematically created through web scraping techniques on the Marktplaats platform, be utilized to analyze and predict market trends in refurbished iPhones? This focuses on the potential of the dataset to inform pricing strategies for sellers based on identified market dynamics.

The sub-question supporting this inquiry is: "How can web scraping be employed to create a comprehensive dataset of refurbished iPhones on the Marktplaats platform?" This addresses the extraction and cleansing of data related to model, storage capacity, and price to ensure its reliability and relevance.

**4. Literature**

**Web scraping**

Using Web Scraping In A Knowledge Environment To Build Ontologies Using Python And Scrapy from:   
<https://www.researchgate.net/profile/Krit-Salah-Ddine/publication/346215371_Using_Web_Scraping_In_A_Knowledge_Environment_To_Build_Ontologies_Using_Python_And_Scrapy/links/5fbcddb892851c933f52508a/Using-Web-Scraping-In-A-Knowledge-Environment-To-Build-Ontologies-Using-Python-And-Scrapy.pdf>

Learning Scrapy from:   
<https://f.lib.dipuni.uz/IT/Learning%20Scrapy.pdf>

**Dynamic pricing / modelling**

Feature-Based Dynamic Pricing from:  
<https://maxccohen.github.io/Feature-Based-Dynamic-Pricing.pdf>

**Sustainability**

Circular Economy of Refurbished Smartphones in the European Union from:  
<https://www.theseus.fi/bitstream/handle/10024/801292/Hapuli_Kumpulainen.pdf?sequence=2>

**Refurbished smartphone market**

Infographic: Global Refurbished Smartphone Market | 2021 from:  
<https://www.counterpointresearch.com/insights/refurbished-smartphone-market-2021-infographic/>

**5. Method**

5.1 Data Collection  
The primary data for this study will be collected through a systematic web scraping process targeting the Marktplaats platform. This process aims to extract listings of refurbished iPhones, focusing on model names, storage capacities, listing prices, and posting dates. Python, utilizing libraries such as BeautifulSoup or Scrapy, will be employed to handle both structured and unstructured web data efficiently. These tools are chosen for their robustness in data extraction, offering the flexibility required to navigate and parse through a diverse range of web page layouts.

5.2 Data Cleansing  
After data collection, a thorough data cleansing procedure will be implemented to ensure the dataset's accuracy and usability. This stage will involve removing duplicate entries, handling missing values, and standardizing entries for model names and storage capacities. Outlier detection algorithms will also be applied to rectify data points that significantly deviate from typical market patterns, potentially indicating data capture errors or anomalous listings. Special attention will be given to exclude abnormal numerical entries, such as placeholder numbers like "123456", which are unlikely to represent genuine prices. Additionally, efforts will be made to identify and remove listings from potentially fraudulent sources, such as those posted by brand new accounts offering items at suspiciously low prices. These measures are intended to enhance the reliability of the dataset by eliminating likely sources of bias and error.

5.3 Data Analysis  
Once a clean dataset is prepared, the analysis will proceed using statistical and data visualization tools in Python, such as pandas for data manipulation and matplotlib for creating visualizations. Initial analysis will use descriptive statistics to outline the basic characteristics of the market. Subsequent analyses will include time series analysis to track price trends over time. This phase will also employ machine learning techniques like regression analysis and clustering to uncover underlying patterns and variables influencing pricing trends in the refurbished iPhone market.

5.4 Utilization of the Dataset  
The insights derived from the data analysis will be structured to highlight how the dataset can be utilized to understand and predict market trends effectively. This will involve discussing the potential applications of the findings in real-world scenarios, such as informing pricing strategies for sellers or guiding inventory decisions based on anticipated market movements. The focus will be on the practical implications of the dataset in enhancing market efficiency and supporting sellers in the secondary market.

5.5 Evaluation  
The evaluation phase will measure the effectiveness of the analytical methods used in this study. It will assess the comprehensiveness of the data captured, the robustness of the cleaning methods, and the accuracy of the analysis in depicting market trends. Feedback loops will be established to continuously refine the data collection and analysis processes based on findings and stakeholder input.

5.6 Ethical Considerations  
Ethical considerations regarding data privacy and the legality of web scraping will be addressed throughout the research. This includes adhering to the terms of service of the Marktplaats platform and ensuring compliance with applicable data protection laws. This ethical framework aims to safeguard personal information and maintain the integrity of the research process.

**6. Planning**

**Week 1:**8 April – 14 April  
Projectconceptualisatie

**Week 2:**15 April – 21 April  
Projectvoorstel goedgekeurd krijgen

**Week 3:**22 April – 28 April  
Opstarten van webscraping om een onbewerkte dataset te verzamelen   
Verslag: scriptiebestand aanmaken met de juiste layout, en titelpagina plus inleiding schrijven

**Week 4:**29 April – 5 Mei  
Dataset opschonen  
Verslag: methodesectie uitwerken

**Week 5:**6 Mei – 12 Mei  
Data visualiseren  
Verslag: Resultaten sectie beginnen

**Week 6:**13 Mei – 19 Mei  
Presentatie

**Week 7:**20 Mei – 26 Mei  
Data analyseren en conclusies

**Week 8:**27 Mei – 2 Juni  
Bespreken van praktische toepassingen zoals prijsstrategieën en voorraadbeheer op basis van markttrends  
Verslag: resultaten en discussie

**Week 9:**3 Juni –9 Juni  
Verslag afmaken en presentatie voorbereiden  
Indien er tijd over is, zullen de webscrapingtechnieken en de opschoningsmethoden die gebruikt zijn voor Marktplaats ook toegepast worden op andere online marktplaatsen zoals Facebook Marketplace, om de resultaten te vergelijken.

**Week 10:**10 Juni – 16 Juni  
Eindpresentatie