**Rough Outline/Working Notes for Webscraping ED Rx Prices**

**Webscraping the variety of costs when self-paying for ED Meds:**

* GoodRxselected over other Rx discount cards because of lower costs, flexibility in dosages, wider range of partner pharmacies, personal success using this card)
* Retail (can get from GoodRx scrape)

**Why do we care:**

* Stats on ED
* Not covered by many health plans… and, when covered, usually at higher co-pay. Always research options under your health plan before assuming self-pay. But, some who have health plan coverage may still prefer to self-pay if they have concerns about their condition being part of their primary medical file.
* The need to remove stigma and fear…
  + Provide stats on ED…
    - roughly increases 10% for every decade of a man’s life… meaning if you are in your 40s, you have a 40% chance of having some level of ED… if you are in your 60s, 60% chance… and so on
    - in other words, ED is extremely COMMON as a man ages… it is to be EXPECTED
    - seeking treatment is becoming more common, but a lot of men are still embarrassed to go to their primary doctor to discuss options
  + Why would I pick this topic for research as a woman?
    - often times, the first person a man will talk to about his ED is his partner
    - if he’s embarrassed, she may be the one who does the research, makes the appointments, and picks up the Rxs
    - women are more used to ‘embarrassing’ medical appointments… to some degree, you could say ‘we’re over it’ when it comes to considering anything embarrassing w/a doctor by the time we’re in our 40s because we’ve already been through decades of gyno exams, breast exams, and for a lot of us, birthing
  + If you search Viagra or Cialis online, you are directed to sites that provide an end-to-end solution, but at a very high cost such as ‘Hims’ and ‘Roman’ and ‘Lemonaidhealth’… all of which provide low cost or free initial consultations, but require that you purchase the meds through them.
  + Mens Health Centers also exist such as ‘Summit Male Medical’. These are often the highest cost solutions, but do provide more extensive consultation, tests, options, etc. But, when ED meds are the recommended treatment, they require that you purchase through them at a very high cost
  + The above two options are legit (costly, but legit). When searching the internet, you will also be directed to many sites that are 100% scams… fake consultation (if any) and fake pills (may want to reference gov’t seizures of fake Viagra, etc. coming in from abroad)
  + Be wary of any healthcare provider that also requires that you obtain the medication from them. This is the model of the free branded razor, but you must buy the branded razor blades at a higher cost… costing you more in the long-term. This business model has been used over and over again very successfully… HP Printers and Print Cartridges… Juul vape and Juul vape pods... etc. IMO, this sort of model should not ever be applied to healthcare… a medical diagnosis should never be made to produce a Rx that the company retains to restrict you from only being able to fill it within their own pharmacy. This creates a conflict of interest for the prescriber working for the company that also owns the pharmacy.
  + There is a much more direct and cost-efficient way to obtain ED meds… through a normal telehealth visit w/prescription sent to a normal US pharmacy. Your telehealth visit may involve a questionnaire and/or phone call and/or video chat (depending on the telehealth provider).
    - If you have insurance, you may have telehealth covered for free under your plan and even though the ED meds may not be covered, the telehealth visit may be free under your plan.
    - Self-pay for a telehealth visit typically costs around $35-$75 (no subscription). ‘Hey Doctor’ charges $39 for a telehealth visit to discuss ED.
  + In the case of GoodRx, they do have a telehealth option ‘HeyDoctor’, but when you use ‘HeyDoctor’, your Rx can be sent to any pharmacy and you choose whether or not to use the GoodRx free Rx card. And, you can use the GoodRx free Rx card whether you went to any regular doctor or ‘HeyDoctor’. There is no required link between the two as in Mens Health Centers and Online Subscriptions such as ‘Hims’ and ‘Romans’.
* etc

**Parameters to put into GoodRx:**

* Zip Code
  + most populated zip codes can be found at this website: <https://localistica.com/usa/zipcodes/most-populated-zipcodes/>
  + may want to pick from this list, but not necessarily the top five… could be a ‘coast-to-coast’ comparison such as…
    - LA, CA 90011
    - Houston, TX 77084
    - Chicago, IL 60629
    - New York, NY 10025
* Rx Name (top 3)
* Rx Dosage (could use low/high, typical/high, or every dosage available)
* Rx # of Pills (could use 30 since this is considered a typical bottle of pills… but, would need to explain that this does not equate to monthly cost since some pills are for daily use and some are taken, as needed)

**GoodRx Website Exploration:**

* Just selecting Rx name from main page:
  + <https://www.goodrx.com/sildenafil> (generic Viagra)
  + <https://www.goodrx.com/tadalafil-cialis> (generic Cialis)
  + <https://www.goodrx.com/vardenafil> (generic Staxyn, Levitra)
* Changing dose to 100mg with 30 pills:
  + <https://www.goodrx.com/sildenafil?dosage=100mg&form=tablet&label_override=sildenafil&quantity=30&sort_type=popularity>
  + … and so on…
* Setting location does not change URL… need to explore this further… where does zipcode show up in source code?
  + Am I going to need to use Selenium so that I can enter a zipcode… what will be the default when spider runs in scrapy?
  + Maybe I just go w/the default and give disclaimer that prices appear to be pretty consistent across the US.
  + Could use most populous zipcode just in NY, NY 10025… this gives a wide variety of pharmacies (that are known across the country) with a wide variety of prices
  + ***NEED to use Selenium if I want to get data by zip code or city, state. Running Scrapy will always assume the search should be done around my geographic location. GoodRx uses javascript to dynamically render pages based on location which is defaulted, but can be changed interactively.***
* All Groups found by:
  + //\*[@id="uat-price-row-coupon-0"]… starting w/0, then 1,2,3…
* Scrapy Items: (example on 4th group… 3 here)
  + Pharmacy Name
    - //\*[@id="uat-price-row-coupon-3"]/div[1]/div/span[2]
  + Retail Price
    - //\*[@id="uat-price-row-coupon-3"]/div[2]/div/span[1]/text()
  + GoodRx Price
    - //\*[@id="uat-price-row-coupon-3"]/div[3]/div[1]/text()

**Other Stuff:**

Trying XPaths…

In Scrapy shell: scrapy shell -s USER\_AGENT="Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/90.0.4430.85 Safari/537.36" "https://www.goodrx.com/sildenafil?dosage=100mg&form=tablet&label\_override=sildenafil&quantity=30&sort\_type=popularity"

response.xpath('//\*[@id="uat-price-row-coupon-0"]/div[1]/div/span[2]/text()').extract()

Out[7]: ['Publix']

**Considerations for Where to Go from Here… WellRx… Selenium…**

The data set is greatly enriched and variation in prices much better represented when showing price differences across locations as well as pharmacies. A lot more analysis can be done… including differences between same pharmacy and different location. Overall, I can expect better impact/interest in my analysis if I include the geographic component.

In order to compare prices by location in GoodRx, I will need to change from Scrapy to Selenium.

* Cons:
  + Changing tools mid-project is a risk that has not always panned out well for me before.
* Pros:
  + Even with searching prices in multiple locations, data set is still *relatively* small so performance should not be an issue.
  + Selenium is a better tool for me to learn long-term because it overlaps into my skillset of automated testing (WinRunner).

If I want to compare prices by location, but want to stick w/Scrapy, I can change GoodRx to WellRx.

* Cons:
  + I have no personal experience w/WellRx and initial analysis on different free Rx card services indicated that GoodRx appears to be the best free Rx card service available. My perception is that GoodRx is the most widely accepted/well known service of this type.
* Pros:
  + Sticking w/the tool I’ve started with and have proven to get the data items I need is safer at this point in the project.

***Conclusion: Time box amount of time to spend on getting webscraping working using Selenium and GoodRx to no more than 6 hours and then re-evaluate.***