



PRICE[®]

ICEAA 2019

“Big Data” Analytics in Operations Research

Presented by:
Cara Cuiule
Grady Noll

Estimate with Confidence™

© 2019 PRICE Systems, L.L.C. All Rights Reserved

Agenda

- Objectives
- Web Scraping and Crawling Definitions
- Methodology
- Limitations
- Future Directions



Objectives

- Create automated processes to collect online data for current or future work
- Validate data to create Cost Estimating Relationships (CERs)

```
339 <div class="text-right">
340 <span
341 class="market-price"><span
342 class="currency">$629.00</span></span><span
343 itemprop="offers" itemscope itemtype="http://schema.org/Offer"><meta
344 itemprop="priceCurrency" content="$" />
345 <span
346 class="product-price-value" itemprop="price" content="499.98"><span
347 class="currency">$</span><span
348 id="product_price">499.98</span></span></span><link
349 itemprop="availability" href="http://schema.org/InStock" />
350 </span></div><div
351 class="right">
```



Name	Price
Charter Arms Undercover Crimson Trace 38 Special Revolver	629
Charter Arms Undercover DAO 38 Special Revolver	383
Charter Arms Undercover DAO Stainless 38 Special Revolver	397
Charter Arms Undercover Green and Black 38 Special Revolver	419.99
Charter Arms Undercover Grey Stainless 38 Special Revolver	424
Charter Arms Undercover Lite 38 Special Revolver	448
Charter Arms Undercover Lite Black and Polished Silver 38 Special Revolver	423.99

Web Scraping vs Web Crawling

- Take data from Hypertext Markup Language (HTML) code
 - Web crawling - Identify webpage structure and links
 - Web scraping - Extract the data

Data scraping	Data Crawling
Involves extracting data from various sources including web	Refers to downloading pages from the web
Can be done at any scale	Mostly done at a large scale
Deduplication is not necessarily a part	Deduplication is an essential part
Needs crawl agent and parser	Needs only crawl agent

<https://www.promptcloud.com/data-scraping-vs-data-crawling/>

Methodology

1. Find websites with relevant data
2. Determine if crawling/scraping from the website is allowed
3. Choose the appropriate tool
4. Obtain and normalize the data
 - Obtain a list of product webpages
 - Get the data from product webpages
5. Application
6. Data validation

Find Websites with Relevant Data

1. Inspect for technical information

- Weight
- Material
- Manufacturer

Condition	NEW
Caliber:	.38 Special
Capacity:	5
Barrel Length:	2 inches
Action:	Double Action / Single Action
Grips:	Lasergrips
Weight (unloaded):	16 oz

Figure: Product specification table on the HyattGunStore.com Website

2. Reverse-engineer cost from price

- Research potential taxes applied
- Collect public company financial data
- Calculate manufacturer profit margin

Legality of Web Crawling/Scraping

*****Potential legal GRAY area*****

- Rules of Thumb:

- Cannot use copyrighted data
- Avoid overburdening company servers
- Do not violate forbidden directories in “Robots Exclusion Protocol” (Robots.txt)

```
User-agent: *  
Disallow: /*?  
Disallow: /account.php  
Disallow: /cart.php  
Disallow: /checkout.php
```

Figure: Example of a Robots.txt

HTML Code Example

- Using the webpage where the data is found, right-click and select “View Page Source”*

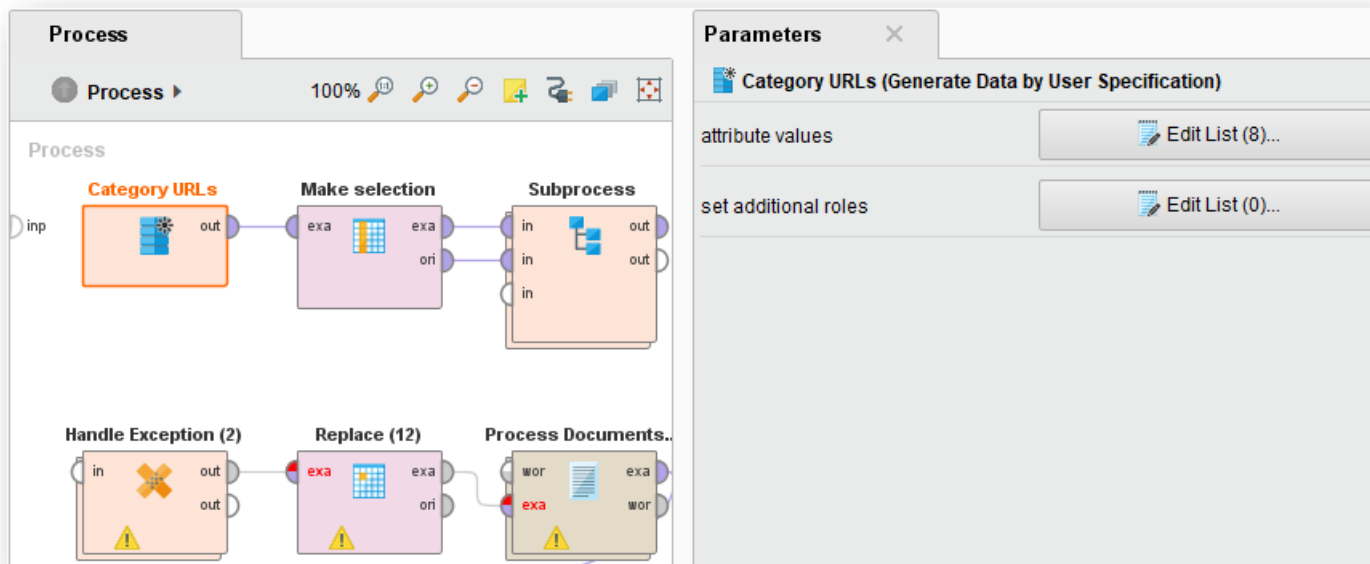
```
339 class="text-right">
340 <span
341 class="market-price"><span
342 class="currency">$629.00</span></span><span
343 itemprop="offers" itemscope itemType="http://schema.org/Offer"><meta
344 itemprop="priceCurrency" content="$" />
345 <span
346 class="product-price-value" itemprop="price" content="499.98"><span
347 class="currency">$<span
348 id="product_price">499.98</span></span></span><link
349 itemprop="availability" href="http://schema.org/InStock" />
350 </span></div><div
351 class="right">
352 <small>
353 </small></div></div></div></div><div
354 class="add-to-cart-buttons text-right"><ul
355 class="button-group radius"><li><button
356 class="button add-to-wishlist" type="button" title="" onclick="javascript: if (FormValidation()) submitForm(document.orderform, 'add2w1', arguments[0]);">
357 <span
358 class="button-right"><span
359 class="button-left"><i
360 class="fa fa-heart"></i></span></span>
361 </button></li><li><button
362 class="button main-button add-to-cart-button simptip-position-bottom simptip-smooth simptip-fade" type="submit" title=" Add to Cart" data-tooltip="Add to Cart">
363 <span
364 class="button-right"><span
365 class="button-left"><i
366 class="fa fa-cart-plus"></i> Add to Cart</span></span>
```

*Can only be done using certain web browsers

Choose the Appropriate Tool, Pt. 1

- RapidMiner

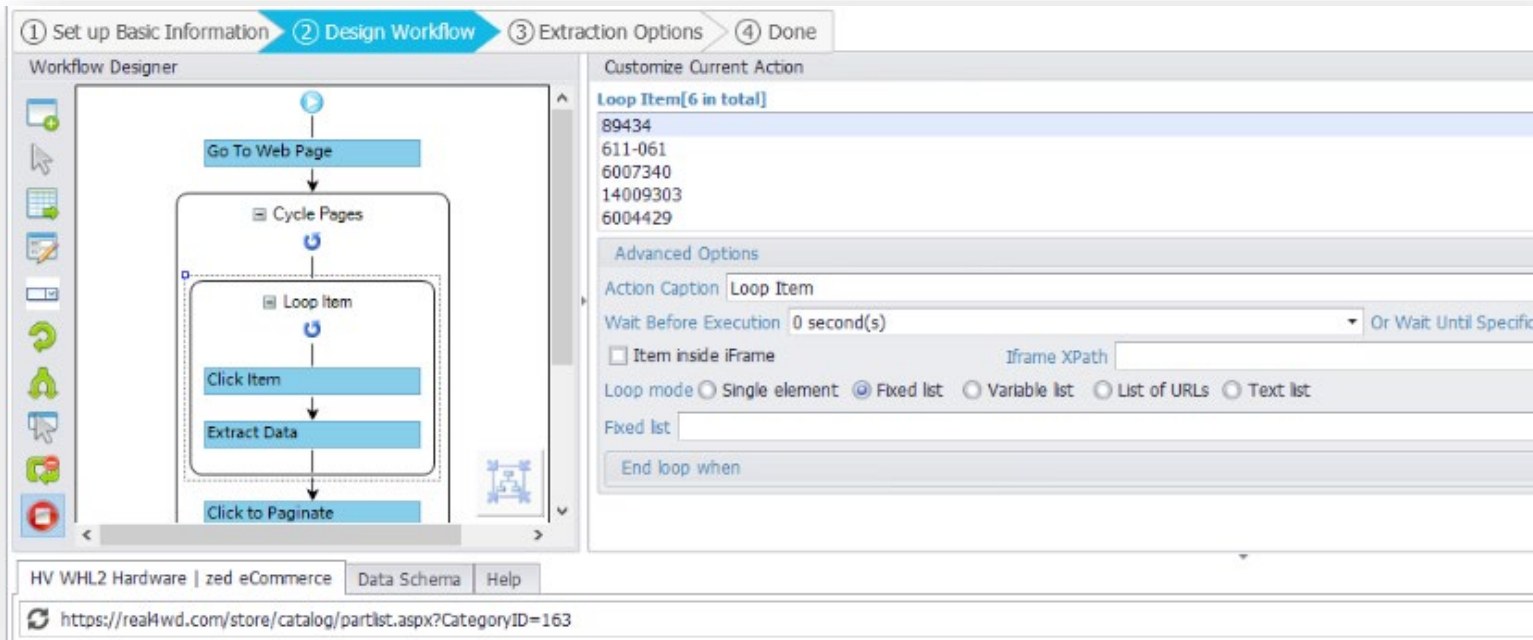
- Customizable operators in a graphical user interface (GUI)
- Web Mining extension
- Transform and validate data
- STEEP learning curve



Choose the Appropriate Tool, Pt. 2

- Octoparse

- Machine learning algorithms in background
- Easy to use “Point and Click”
- Limited application



Workflow Designer

① Set up Basic Information ② Design Workflow ③ Extraction Options ④ Done

Workflow Designer

Go To Web Page

Cycle Pages

Loop Item

Click Item

Extract Data

Click to Paginate

Customize Current Action

Loop Item[6 in total]

89434
611-061
6007340
14009303
6004429

Advanced Options

Action Caption Loop Item

Wait Before Execution 0 second(s) Or Wait Until Specific

Item inside iFrame iFrame XPath

Loop mode Single element Fixed list Variable list List of URLs Text list

Fixed list

End loop when

HV WHL2 Hardware | zed eCommerce Data Schema Help

<https://real4wd.com/store/catalog/partlist.aspx?CategoryID=163>

Obtain and Normalize the Data

- Collect product webpage
- Extract data with “string matching”

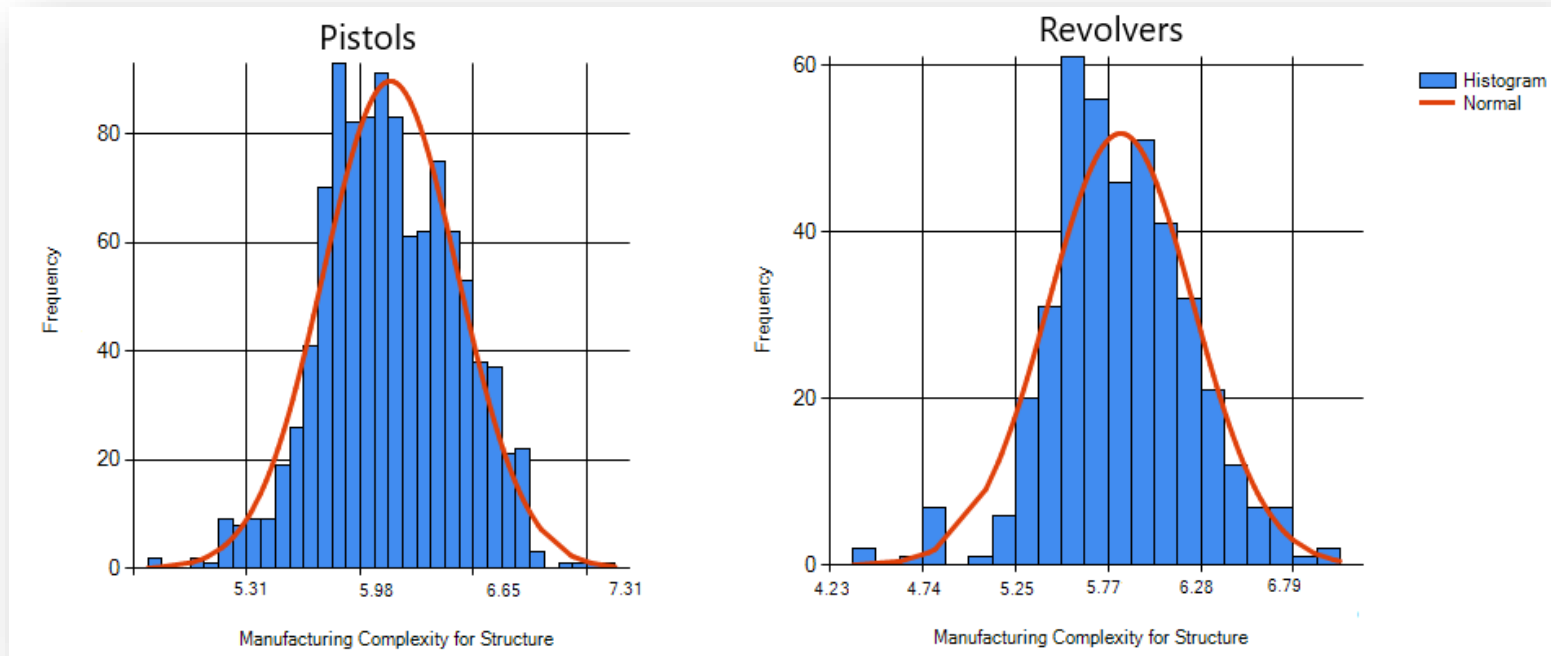
```
class="market-price"><span
class="currency">$629.00</span></span><span
itemprop="offers" itemscope itemtype="http://schema.org/Offer"><meta
itemprop="priceCurrency" content="$" />
```

- Sanitize into data set

Name	Price	Item SKU	Brand	Caliber	Capacity	Condition	Frame Construction
Charter Arms Undercover Crimson Trace 38 Special Revolver	629	73824	Charter Arms	.38 Special	5	NEW	Steel
Charter Arms Undercover DAO 38 Special Revolver	383	13811	Charter Arms	.38 Special	5	NEW	Steel
Charter Arms Undercover DAO Stainless 38 Special Revolver	397	73811	Charter Arms	.38 Special	5	NEW	Steel
Charter Arms Undercover Green and Black 38 Special Revolver	419.99	23820	Charter Arms	.38 Special	5	NEW	Steel
Charter Arms Undercover Grey Stainless 38 Special Revolver	424	43820	Charter Arms	.38 Special	5	NEW	Steel
Charter Arms Undercover Lite 38 Special Revolver	448	53820	Charter Arms	.38 Special	5	NEW	Alloy

Application

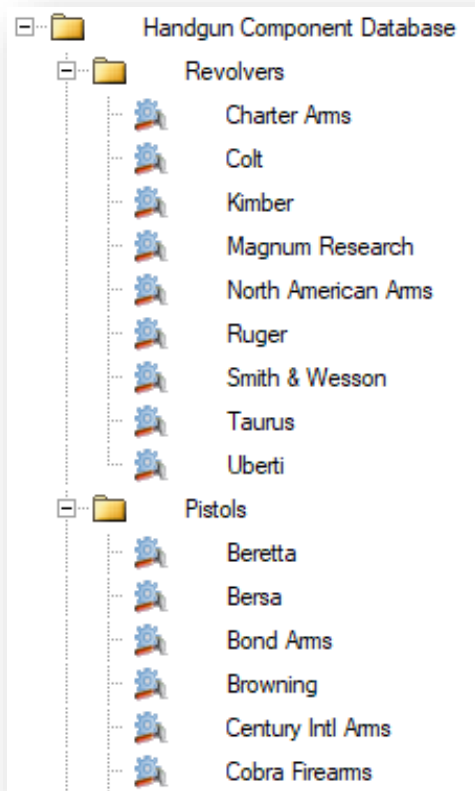
- Calibrate Manufacturing Complexity (cost/weight unit)
- Test dataset for potential errors



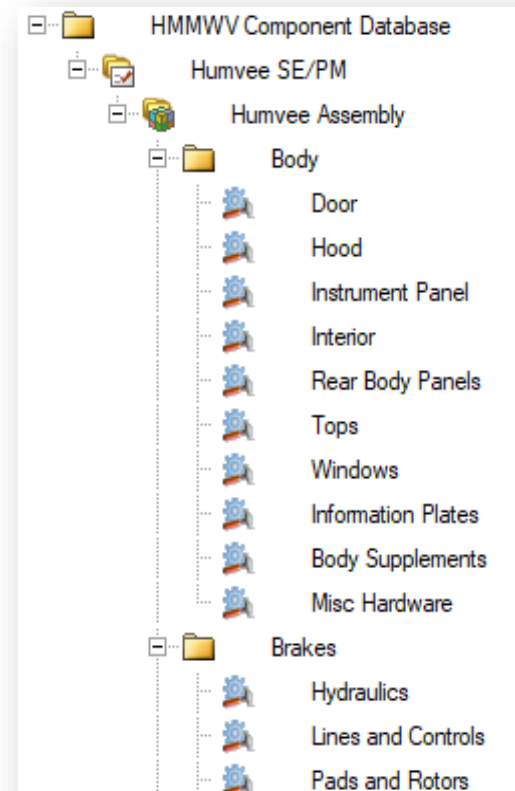
Database Creation

- Two types of databases

Item Listings



Template Components



Validation

- Compare calibrated data to real data
- Goal – Be within one standard deviation of error (+/-33%) of actuals
- Examples of different types of estimates:
 - Early program estimates
 - Price-to-win bidding



Data Validation, Case 1 - Handguns

- **XM17 Modular Handgun System Competition**
 - Contract - \$5.4 million for about 7,100 pistols
 - Seven firearms in separate bids
 - On average, estimates 10% below actuals

- **2005 Sig Sauer and Heckler, Koch Handgun contracts**
 - Two contracts, 65,000 pistols each
 - Valued at \$23.7 and \$26.2 million
 - Caliber of 9mm, 40 S&W, 0.357 Sig
 - Our estimate - \$25.6 million

Data Validation, Case 2 - Humvee

- Created PBS based on assembly designations
- Assigned new QTY Per Next Higher Level to needed items
- System Results:
 - Basic Humvee Weight: 5,200 lbs
 - Humvee PBS Weight: 5,041 lbs
 - Basic Humvee Price: \$70,000
 - Humvee PBS Price: \$74,207



Other Types of Databases

- **IT Hardware**

- Workstations, Laptops
- Servers, Data storage
- Printers
- Routers



- **Long Guns**

- Rifles (Sniper and Tactical)
- Shotguns



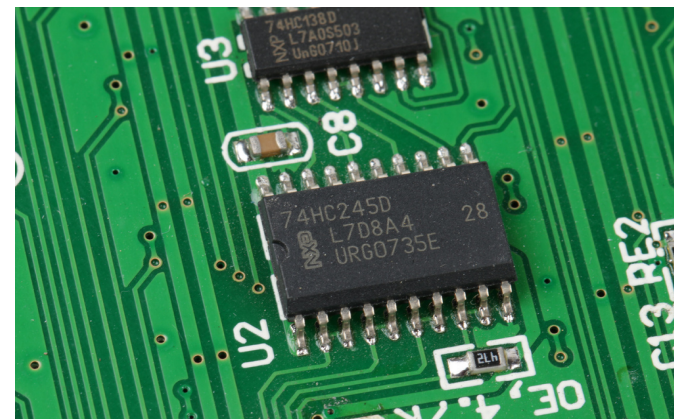
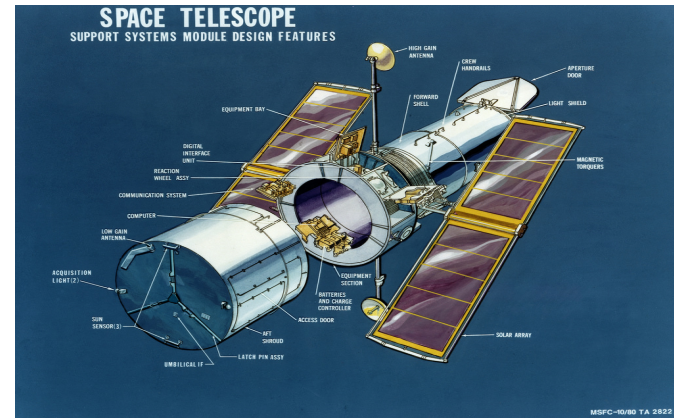
Limitations

- Interpolated dataset values
 - Production Numbers, Manufacturing processes, etc.
- Retailer profit margin
- Structure vs Electronic distribution
- Lack of API integration



Future Directions

- More databases influenced by demand
 - Cloud Service Pricing
 - Microcircuits
 - Space Components
 - Avionics Suites
 - Armaments



Questions?



cara.cuiule@pricesystems.com

grady.noll@pricesystems.com

Backup

Calculating Manufacturer Profit Margin

		<i>Value</i>	
<u>Revenue</u>	903		
<u>Cost of Revenue</u>	528		
<i>Direct Cost</i>	230		
<i>Overhead Cost</i>	298	130	Overhead
<u>General and Administrative</u>	175	33.2%	
Operating G&A	175		
<i>Research Development</i>	10		
<i>Selling General and Administrative</i>	137		
<i>Non Recurring</i>			
<i>Others</i>	28		
Other G&A	-		
<u>Total Loaded Cost</u>	703		
<u>Net Income</u>	128	18.2%	Fee or Profit

XM17 Modular Handgun System



Competition – Results Table

Brand	Firearm	TP Estimate	%Error
Beretta	M9A3	\$7,471,667	38.4%
Beretta	APX	\$3,879,026	-28.2%
CZ-USA	P-09	\$4,066,033	-24.7%
FN USA	509	\$4,438,781	-17.8%
Glock	19	\$4,675,500	-13.4%
Sig Sauer	P320	\$4,914,740	-9.0%
Smith & Wesson	M&P 9	\$4,564,371	-15.5%
AVERAGE		\$4,858,588	-10.0%