**MotoSports.com**

**Estimated time to complete the site: 2 days**

**Challenges:** Need to investigate more to get the product details by size because from details page we need to make post call to get product details by size and all the post parameters are not available on the page.

Scraping OEM Parts and Tires: Site page structure is different for OEM parts and Tires

Need to check if both can be added under single site otherwise need to add two different sites for tires and OEM parts.

Tires 🡺 Need more investigation for getting product price by size

1. Landing Page : Home page

url: <http://www.motosport.com/>

Page Content: Page contains category links of Different type motorcycle

Capture motorcycle type links

Page Type: Category Page

1. Individual motorcycle link

url: <http://www.motosport.com/GB/motorcycle>

this url would be captured from the home page html

Page Type: Category Page

Page content: Page contains category links for OEM, Tires etc.

Capture Tires links

1. Individual Tyre link

url: <http://www.motosport.com/GB/motorcycle/tire-combos>

Page Content: page contains links of product by brand, style and price (filters) and displays list of products

Page Type: list page

Capture product detail link

1. Individual product details link

UIr: <http://www.motosport.com/gb/metzeler-enduro-3-sahara-tire-combo>

Page content: Page contains product details with default price (lowest) page also contains dropdown list to get product price by size.

\*\* To get product price by size need to make post call for each list

Example url for one product size:

Url : <http://www.motosport.com/ajax/pdp>

**Post Parameter:**

action=dropdown&skus={"Front-Tire":"MTZ001C-X001-Y001","Rear-Tire":""}&clickedsku=MTZ001C-X001-Y001&mainsku=MTZ001K-X001-Y001&producttype=config&prod\_uid=1325096&quantity=1

**Post parameters minimization**:

action=dropdown&skus={"Front-Tire":"","Rear-Tire":"MTZ001J-X001-Y002"}

mainsku=MTZ001N-X001-Y001

producttype=config

prod\_uid=1325099&quantity=1

**Mandatory Header:**

Content-Type: application/x-www-form-urlencoded; charset=UTF-8

Cookie(constent): Cookie: PHPSESSID=ni737pcddug81vsthfcv2vfga1;

Request Headers:

Content-Length: 172

User-Agent: Mozilla/5.0 (Windows NT 6.1; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/39.0.2171.95 Safari/537.36

Content-Type: application/x-www-form-urlencoded; charset=UTF-8

Accept: application/json, text/javascript, \*/\*; q=0.01

Accept-Encoding: gzip, deflate

Accept-Language: en-US,en;q=0.8,it;q=0.6

Cookie: PHPSESSID=ni737pcddug81vsthfcv2vfga1; CustomerId=491341688+24b704f7eee9e26b436f3f5d2b923996; RES\_TRACKINGID=508058092556893; wlcme=true; \_\_utmt=1; s\_cc=true; s\_fid=30C36DCA1BD7318B-3502AA2DCA551BFF; s\_sq=%5B%5BB%5D%5D; RES\_SESSIONID=986162528861314; ResonanceSegment=1; 90516023-VID=112730536825017; 90516023-SKEY=5344439268793212137; HumanClickSiteContainerID\_90516023=STANDALONE; msoSkin=motorcycle; \_\_utma=1.635204522.1421407394.1421407394.1421407394.1; \_\_utmb=1.60.9.1421410473224; \_\_utmc=1; \_\_utmz=1.1421407394.1.1.utmcsr=(direct)|utmccn=(direct)|utmcmd=(none); \_\_utmv=1.|1=customer=NONCUSTOMER=1

Host: www.motosport.com

Cookie minimization:

Cookie: PHPSESSID=ni737pcddug81vsthfcv2vfga1;

PHPSESSID=ni737pcddug81vsthfcv2vfga1

Need to pass cookie:

Fields available:

Title:

Price:

Product size:

Product site Item id: ("Metzeler-Enduro-3-Sahara-Tire-Combo")

Manf. Id: not available

Availability:

Price range

OEM 🡺

1. Landing Page : Home page

url: <http://www.motosport.com/>

Page Content: Page contains category links of Different type motorcycle

Capture motorcycle type links

Page Type: Category Page

1. Individual motorcycle type link

url: <http://www.motosport.com/GB/dirtbike>

this url would be captured from the home page html

Page Type: Category Page

Page content: Page contains category links for OEM, Tires etc.

Capture OEM parts links for all Brands

1. Individual OEM brand link

Url: <http://www.motosport.com/dirtbike/oem-parts/honda>

This link would be captured from (SL 2) page.

Page Type: category page

Page content: Page contains links of Model of the brand (Honda)

Capture links of all model to get all the parts links

1. Individual model link

Url: <http://www.motosport.com/gb/dirtbike/oem-parts/honda/2015/crf110f>

This link would be captured from the OEM Brank link page (SL 3)

Page Type: Category Page

Page content: Page contains Parts categories

Capture links of parts categories

1. List of parts of a category

Url: <http://www.motosport.com/gb/dirtbike/oem-parts/honda/2015/crf110f/air-cleaner>

This url would be captured from Individual Model link (SL 4)

Page content: Page contains List of parts of one category

Page Type: List page

Capture list of products

1. Detail page : Individual parts details

Url: <http://www.motosport.com/dirtbike/oem-parts/honda/2015/crf110f/air-cleaner/hose%2c-breather/15761-kyk-910>

This url would be capture from the List page (SL 6)

Page Type: Detail page

Page content: Page contains details of a parts

Capture product details

Fields Available:

SKU:

Title:

Description:

Retail Price:

Our Price:

Stock status:

Previous price:

Delivery Description:

**Issues of scraping the site:**

**Tires** 🡺

1. Details page contains list of items (items by product size/model)
2. Different products details page format is different, like some of them has 2 dropdown list where all the content of these dropdown list are available on the page and some of the product has 3 dropdown list where we choose one list it loads another list. After that when year, make and model are chosen it loads page where another dropdown list appears with product sizes, but on this page both product information available don’t need to make another call.

Proposed Solution: TO DO