



WEB SCRAPPING

Motor Bekas OLX Indonesia

*Teuku Muhammad Roy Adrian
2008107010069*



01. REQUIREMENTS
for scraping OLX

02. INITIALIZE PROJECT

03. SCRAPING

04. RESULT

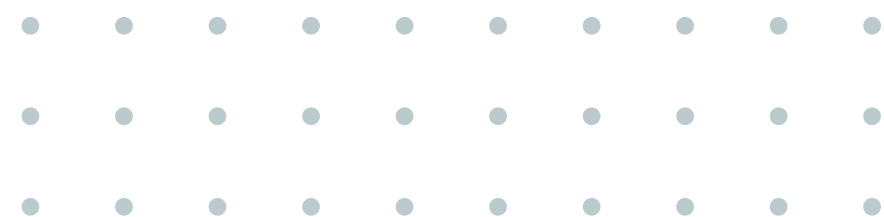


TABLE OF CONTENT



01.

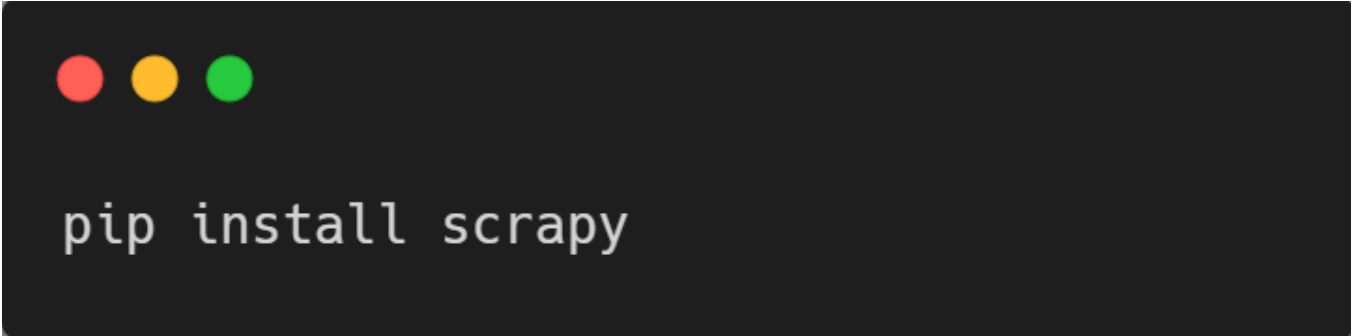
REQUIREMENT



- 
- **Python**
 - **Virtual Environment**
 - **pip**
 - **scrapy**
 - **text editor**

How to install scrapy

After activate virtual environment, use this command



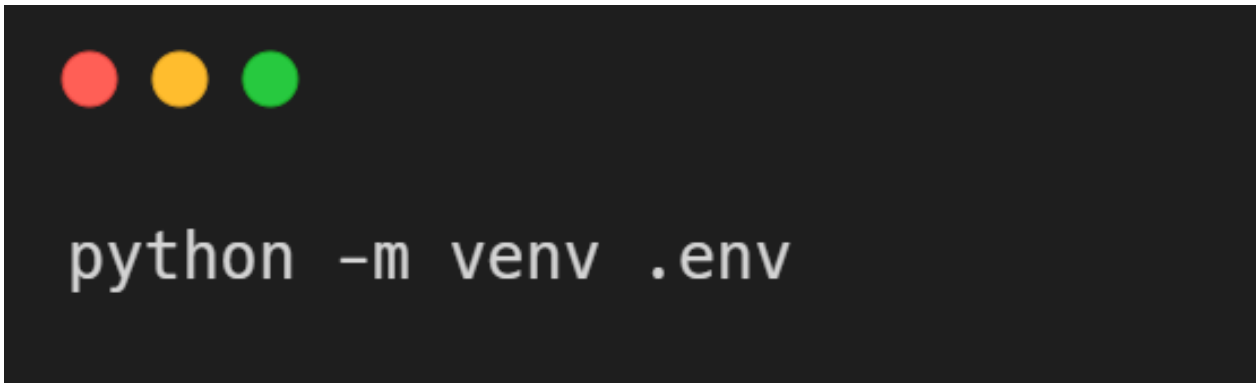
```
pip install scrapy
```

How to make Virtual Environment

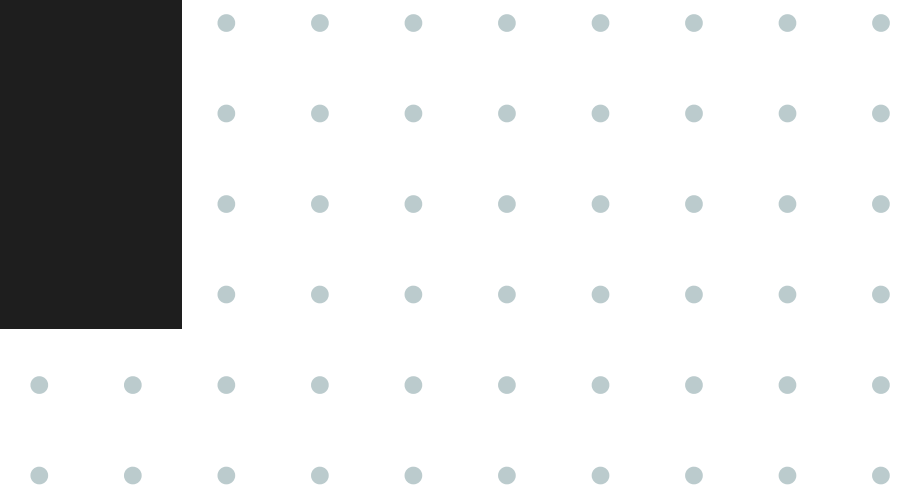


```
python3 -m venv .env
```

OR

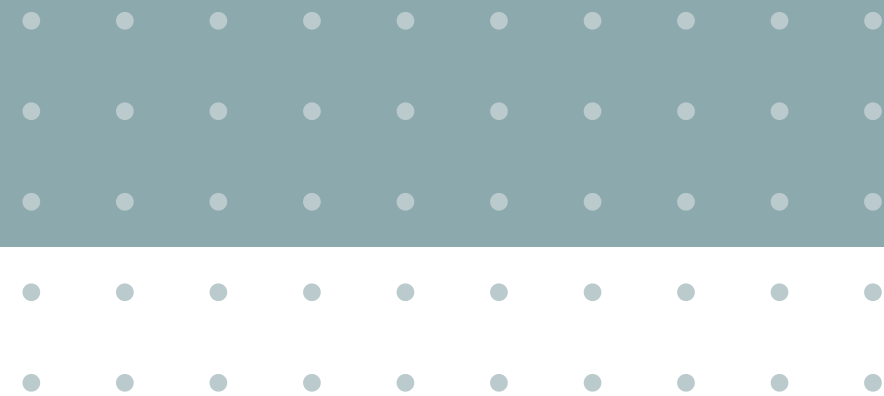


```
python -m venv .env
```



02.

INITIALIZE PROJECT

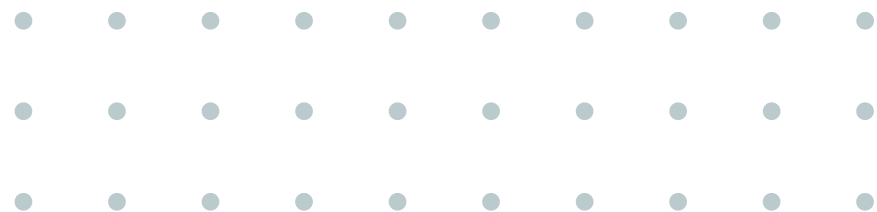


Create scrapy project

Setelah proses instalasi, gunakan perintah **scrapy startproject <namaproject>** untuk membuat project scraping menggunakan scrapy



```
scrapy startproject olxmotor
```



03.

SCRAPING



Scraping

Setelah proses membuat project, jalankan perintah **scrapy genspider [nama_file] [link]** untuk membuat file scraping python secara otomatis di directory spiders

Disini saya melakukan crawling data motor bekas pada website OLX dari page 1 hingga 30 dan juga mengambil class element menggunakan inspect element pada browser untuk mengambil item harga, tahun, dan brand motor.

```
import scrapy

class OlxmotorScrapingSpider(scrapy.Spider):
    name = "olxmotor_scraping"
    allowed_domains = ["www.olx.co.id"]

    def start_requests(self):
        starts_url = "https://www.olx.co.id/motor-bekas_c200?page={}"
        for page_number in range(31): #looping page 1 to 30
            url = starts_url.format(page_number)
            yield scrapy.Request(url, self.parse)

    def parse(self, response):
        price = response.css('._2Ks63::text').extract()
        year = response.css('._YBbhy::text').extract()
        brand = response.css('._2poNJ::text').extract()

        for item in zip(price, year, brand):
            items_info = {
                'price': item[0],
                'year': item[1],
                'brand': item[2]
            }
            yield items_info
```


Scraping

Proses scrap dapat dimulai dengan menggunakan perintah **scrapy crawl [nama_file]**. Disini saya menggunakan file **olxmotor_scraping**. Saya menjalankan perintah **scrapy crawl olxmotor_scraping -o output.json** yang dimana output hasil scrap akan disimpan dalam format .json.

```
2023-11-09 02:50:12 [scrapy.extensions.feedexport] INFO: Stored json feed (581 items) in: output.json
2023-11-09 02:50:12 [scrapy.statscollectors] INFO: Dumping Scrapy stats:
{'downloader/request_bytes': 19562,
 'downloader/request_count': 32,
 'downloader/request_method_count/GET': 32,
 'downloader/response_bytes': 3617933,
 'downloader/response_count': 32,
 'downloader/response_status_count/200': 31,
 'downloader/response_status_count/404': 1,
 'elapsed_time_seconds': 3.925354,
 'feedexport/success_count/FileFeedStorage': 1,
 'finish_reason': 'finished',
 'finish_time': datetime.datetime(2023, 11, 8, 19, 50, 12, 674167, tzinfo=datetime.timezone.utc),
 'httpcompression/response_bytes': 18584790,
 'httpcompression/response_count': 32,
 'httperror/response_ignored_count': 1,
 'httperror/response_ignored_status_count/404': 1,
 'item_scraped_count': 581,
 'log_count/DEBUG': 616,
 'log_count/INFO': 12,
 'response_received_count': 32,
 'robotstxt/request_count': 1,
 'robotstxt/response_count': 1,
 'robotstxt/response_status_count/200': 1,
 'scheduler/dequeued': 31,
 'scheduler/dequeued/memory': 31,
 'scheduler/enqueued': 31,
 'scheduler/enqueued/memory': 31,
 'start_time': datetime.datetime(2023, 11, 8, 19, 50, 8, 748813, tzinfo=datetime.timezone.utc)}
2023-11-09 02:50:12 [scrapy.core.engine] INFO: Spider closed (finished)

(.env) C:\Programming\Scraping\olxscrap\olxmotor\olxmotor\spiders>
```

04.

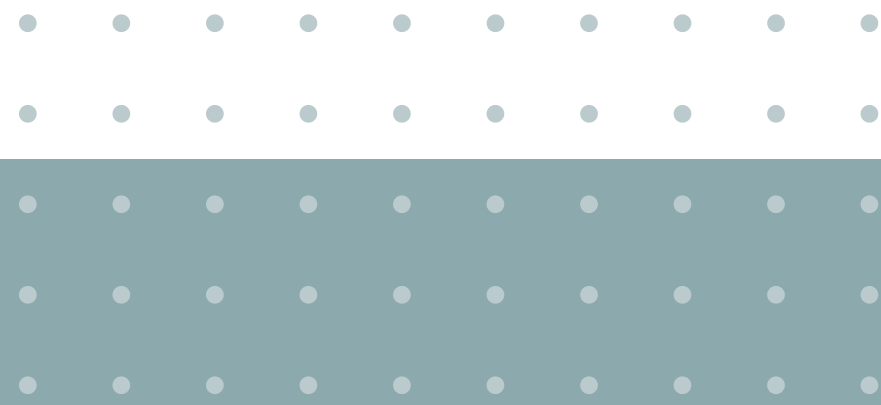
RESULT



Result

Berikut merupakan data yang diperoleh yang menampilkan ratusan baris data price, year, dan brand motor bekas dari website OLX.

```
548 {"price": "Rp 6.900.000", "year": "2013", "brand": "Soul GT tahun 2013 plat B DKI"},
549 {"price": "Rp 8.800.000", "year": "2016", "brand": "Fino fi tahun 2016 surat lengkap"},
550 {"price": "Rp 17.500.000", "year": "2010", "brand": "PUTRA MOTOR 108"},
551 {"price": "Rp 9.900.000", "year": "1988", "brand": "Vespa px strada aom"},
552 {"price": "Rp 6.500.000", "year": "2017", "brand": "Mio M3 125 tahun 2017 surat BPKB only plat B DKI"},
553 {"price": "Rp 25.500.000", "year": "2013", "brand": "PUTRA MOTOR 108"},
554 {"price": "Rp 21.500.000", "year": "2019", "brand": "PUTRA MOTOR 108"},
555 {"price": "Rp 7.900.000", "year": "2017", "brand": "Vixion advance tahun 2017 surat lengkap"},
556 {"price": "Rp 9.300.000", "year": "2016", "brand": "Beat fi tahun 2016 surat lengkap pajak hidup"},
557 {"price": "Rp 15.500.000", "year": "2018", "brand": "PUTRA MOTOR 108"},
558 {"price": "Rp 11.500.000", "year": "2017", "brand": "PUTRA MOTOR 108"},
559 {"price": "Rp 7.900.000", "year": "2014", "brand": "Beat fi tahun 2014 surat lengkap pajak hidup"},
560 {"price": "Rp 15.800.000", "year": "2021", "brand": "Vario 125 cbs merah"},
561 {"price": "Rp 7.800.000", "year": "2012", "brand": "PUTRA MOTOR 108"},
562 {"price": "Rp 6.600.000", "year": "2010", "brand": "Vario karbu 110 tahun 2010 plat B DKI"},
563 {"price": "Rp 18.000.000", "year": "2013", "brand": "Kawasaki Ninja RR 2013 Asli Kuning"},
564 {"price": "Rp 21.200.000", "year": "2022", "brand": "[GARANSI RANGKA SE-UMUR HIDUP] Cash/Kredit Honda New Vario 160 Cbs"},
565 {"price": "Rp 14.500.000", "year": "2018", "brand": "Yamaha Lexi 2018"},
566 {"price": "Rp 42.550.123", "year": "2019", "brand": "VESPA SPRINT S IGET ABS FACELIFT 2019 GREY AVIO MURAH"},
567 {"price": "Rp 18.900.000", "year": "2018", "brand": "FULL CARBON AEROX 2018 VVA 155 ISTIMEWA R15 ADV NMAX BEAT VARIO PCX"},
568 {"price": "Rp 10.555.000", "year": "2012", "brand": "Honda Vario 125 2012 Lengkap smuaON mesinOK noPR jual cepat TT pcx BU"},
569 {"price": "Rp 14.850.000", "year": "2016", "brand": "Honda CBR 2016 REPSOL HARGA NETT"},
570 {"price": "Rp 375.000.000", "year": "2023", "brand": "2024 All New ZX6R ZX636 brand new ready"},
571 {"price": "Rp 55.000.000", "year": "2022", "brand": "Vespa Primavera Putih 2022 Upgrade shock ohlins dan ban pirelli diablo"},
572 {"price": "Rp 20.500.000", "year": "2021", "brand": "YAMAHA R15 V3 2021 HARGA NETT Ninja,cbr,r25,gsx,nmax,aerox,vario,mt15"},
573 {"price": "Rp 340.000.000", "year": "2013", "brand": "BMW GS 1200 K25"},
574 {"price": "Rp 10.500.000", "year": "2017", "brand": "HONDA SUPRA GTR 150 2017 PAJAK HIDUP HARGA NETT"},
575 {"price": "Rp 60.000.000", "year": "2017", "brand": "Vespa Sprint Iget 2017 Full Mod"},
576 {"price": "Rp 26.800.000", "year": "2020", "brand": "Dijual honda pcx 2020 Polar white ISTIMEWA"},
577 {"price": "Rp 21.500.000", "year": "2022", "brand": "Nozomi 3 roda 2022 km rendah hg 21.5 bisa TT motor"},
578 {"price": "Rp 7.900.000", "year": "2014", "brand": "MIO GT FI TH 2014 CASH/ TT/ CC BUNGA0%/KREDIVO/ SPAY/ QRIS"},
579 {"price": "Rp 28.500.000", "year": "2013", "brand": "SUZUKI INAZUMA GW 250 2013 HARGA NETT"},
580 {"price": "Rp 15.750.000", "year": "2017", "brand": "Yamaha aerox 2017 HARGA NETT"},
581 {"price": "Rp 105.500.000", "year": "2021", "brand": "[KM 800 PERAK] ZX25R 2021 ABS KRT LIKE NEW ZX25 ZX 25 NINJA R25 CBR"},
582 {"price": "Rp 27.000.000", "year": "2018", "brand": "Jual Motor Honda Vario 2018 LED Full Mods"}
583 ]
```



THANK YOU

