

Shedding Light on the Darknet

Anna Bladey, WanQi Tay,
Yingkun Zhu, Kunal Shukla
May 31, 2018

Agenda

- Executive Summary
- Business Use Case
- Methodology
 - Tools
 - Data Sources
 - ETL Process
 - Dimensional Model
- Findings
- Recommendations
- Lessons Learned
- Appendix

Executive Summary

- Analyzed data from darknet market webscrapes and economic metrics in order to better understand internet crime
- Information will supplement existing Law Enforcement intelligence and aid in efficient allocation of LE resources
- Data management and analysis was conducted using open source software



Business Use Case

- Law Enforcement agencies are working to combat cybercrime (online trafficking of contraband, data breaches, identity theft)
- Agencies depend on up-to-date, detailed intelligence in order to efficiently allocate limited resources
- Insights generated from darknet data may assist Law Enforcement with preventing and eliminating cybercrime



Tools

- Amazon Web Services (Relational Database Service) - Hosting



- MySQL (Workbench) – Database Design & Implementation



- Data Cleaning

- Excel – Data Preprocessing
- R – Cluster Analysis



- Tableau - Visualizations / Insights



Data Sources

Name	Description	Num. records
Hansa market listings	Market data web scrape of item listings from Hansa market, December 2016	14,773
Valhalla market listings	Market data web scrape of item listings from Valhalla market, October 2016	16,511
Agora market listings	Market data web scrape of item listings from Agora market, 2014-2015	109,621
Country GDP	World Bank and OECD National Accounts data for 2014-2016	264
Country Population	World Bank and OECD National Accounts data files for 2014-2016	264



ETL - fact_listings

Consolidated listing-level data from Agora, Hansa, and Valhalla webscrapes into one fact_listings table

- Dropped columns that were not shared among all subtables
- Created fact_listings.market to track data origin
- Standardized all currencies to USD
- Cleaned fact_listings.shipFromCountry:
 - Clustered related abbreviations and polymorphisms (e.g. “U.S.A.”, “US”, “UnitedStates” = “United States”
 - shipFromCountry = multiple locations or a region were converted to new category “multiple”
 - shipFromCountry = NA or ‘Undeclared’ were converted to blank
- Fact_listings.category: Reduced the number of categories by clustering related ones together

100% C 1.2										Result Grid										Filter Rows: Search										Edit: Export/Import: Fetch rows: 10									
listingID	title	description										categoryID	priceUSD	vendor	shipFromCountryID	marketplace																							
88440	0.5g Fishscale Cocaine - 85-90% Purify - AUS	FE ONLY FISHSCALE COCAINE 0.5g of Fishscale Cocaine 85-90% Purify V...										23	156.04	AuxTech	12	Agora																							
88441	250 Grams of Premium Fishscale Cocaine HQ	Address MUST contain the following info: Name/surname!!! Street-name!!! Ph...										23	19953.61	MailMaker	250	Agora																							
88442	0.5g PREMIUM UNCUT COCAINE 80%	Bolivian origin, 80% Product. This is a very clean product and it has a very c...										23	87.57	MegaMindy	80	Agora																							
88443	50g Benzocaine - Pharmaceutical Grade	Benzocaine - Pharmaceutical Grade Vacuum packed. The best cut for cocaine!										23	35.23	The source	50	Agora																							
88444	Copy of 1 x 1G Straight Off The Brick x3	I am now officially selling a much more potent product for slightly higher th...										23	52.30	TheLeadLady	1	Agora																							
88445	10g Benzocaine Pharmaceutical Grade	Benzocaine - Pharmaceutical Grade Vacuum packed. The best cut for cocaine!										23	11.78	the_source	80	Agora																							
88446	0.5g Pure Fishscale Cocaine AUS CUSTOM P...	We came across some high quality cocaine for a real good price and wanted t...										23	43.63	Nevila	0.5g	Agora																							
88447	1000UF Pure Fish Scale Cocaine	Reviews - http://alexcocaine1794pc.onindex.gr/p/1956-2350/megastuff794pc...										23	7152.44	Medusa - SFR	1000	Agora																							
88448	7g High Quality Cocaine UK Vendor	This will be our listing for 1g of our top quality cocaine. We will always upda...										23	68.84	technology80	80	Agora																							
88449	18grams - Raw Cocaine 76%+ Products4you	***** READ OUR PROFILE: TERMS OF CONDITION - Produ...										23	124.75	Products4You	57	Agora																							
88450	Hq Pure Cocaine Uncut 10 g	Pure Uncut Fishscale Cocaine - What you see is what you get!!! http://an...										23	962.01	TheLithianMaster	115	Agora																							
88451	Custom Kyp											23	11.52	Medusa	100	Agora																							
88452	283g PREMIUM UNCUT COCAINE 80%	Bolivian origin, 80% Product. This is a very clean product and it has a v...										23	2394.76	MegaMindy	80	Agora																							
88453	1 gram 70% Medium-High Grade Cocaine	PLEASE READ MY PROFILE & TERMS PAGE BEFORE ORDERING. This...										23	6.03	DOPECARTEL	250	Agora																							
88454	662g66k66 3.0 gram pure uncut cocaine 90%...	!Our goal is to provide the highest quality cut on the market, and three gr...										23	304.25	UIDPrice	115	Agora																							
88455	Hq Pure Cocaine Uncut 10g	Pure Uncut Fishscale Cocaine - What you see is what you get!!! http://an...										23	7846.77	TheLithianMaster	115	Agora																							
88456	Hq Pure Cocaine Uncut 0.5 g	Pure Uncut Fishscale Cocaine - What you see is what you get!!! http://an...										23	72.62	TheLeadLady	115	Agora																							
88457	Custom order Peruvian	1x1u Kiwi100										23	317.96	GermanTrustSupply	115	Agora																							
88458	100g McGrinlock Cocaine 90%+ Pure	100 grams cocaine - Shipped to Canada, the US! and the UK only! Our coke...										23	4778.37	McGrinlock	266	Agora																							
88459	662g66k66 4.0 gram pure uncut cocaine 90%...	!Our goal is to provide the highest quality cut on the market, and three gr...										23	390.23	UIDPrice	115	Agora																							
88460	COCAINE 1.7grams	High Quality Cocaine Hydrochloride 1.7gram net weight Straight from the Ki...										23	167.51	certificates24	34	Agora																							
88461	100g Peruvian Cocaine Hq 90% uncut 84-...	o t g - o																																					

ETL - dim_categories & dim_metacategories

- snowflaked dim_category out of facts table
 - 59 categories
- Created metacategories to further simplify category clusters for ease of future analysis, and snowflaked dim_metacategories out of dim_category
 - 13 metacategories

1
2 • SELECT * FROM dim_category;

100% 27:2

Result Grid Filter Rows: Q Search

category_ID	category_name	metacategory_ID
1	Chemicals	112
2	Counterfeits/Accessories	100
3	Counterfeits/Clothing	100
4	Counterfeits/Electronics	100
5	Counterfeits/Money	100
6	Counterfeits/Unclassified	100
7	Counterfeits/Watches	100
8	Data/Accounts	112
9	Data/Pirated	112
10	Data/Software	112
11	Digital Goods	112
12	Drug Paraphernalia	112
13	Drugs/Barbiturates	101
14	Drugs/Benzos	101
15	Drugs/Cannabis	101
16	Drugs/Dissociatives	101
17	Drugs/Ecstasy	101
18	Drugs/Opioids	101
19	Drugs/Prescription	101
20	Drugs/Psychedelics	101
21	Drugs/R/Cs	101
22	Drugs/Steroids	101
23	Drugs/Stimulants	101
24	Drugs/Unclassified	101
25	Drugs/Weight loss	101
26	Electronics	102
27	Erotica	112
28	Forgeries/Unclassified	103
29	Fraud Related	104
30	Guides & Tutorials	112
31	Info/eBooks/Allens/UFOs	106

1
2 • SELECT * FROM dim_metacategory;

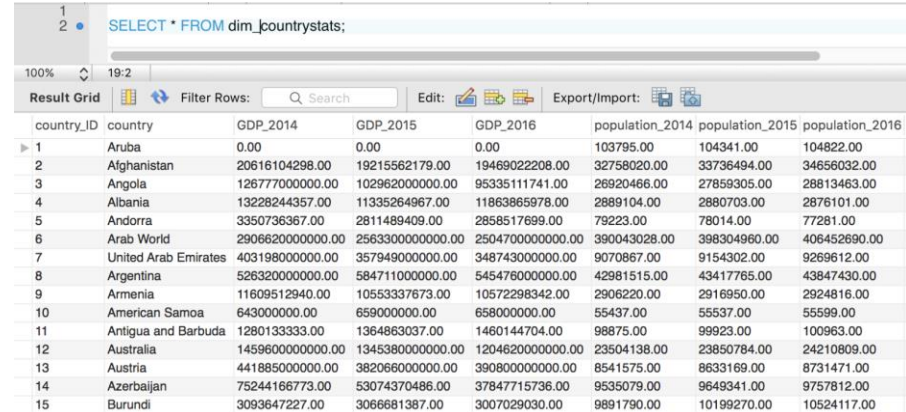
100% 32:2

Result Grid Filter Rows: Q Search

metacategory_ID	metacategory_name
100	Counterfeits
101	Drugs
102	Electronics
103	Forgeries
104	Fraud Related
105	Hacked Accounts
106	Info/eBooks/Guides/Tutorials
107	Jewelry
108	Pirated
109	Reagents/Lab Chemicals
110	Services
111	Weapons
112	Misc
NULL	NULL

ETL - dim_countrystats

- Created dim_countrystats
 - Combined GDP and population data from years 2014-2016 (years of market data scrapes)
 - Standardized country names in fact_listings to be consistent with dim_countrystats naming conventions (e.g. “Russian Federation” rather than “Russia”)

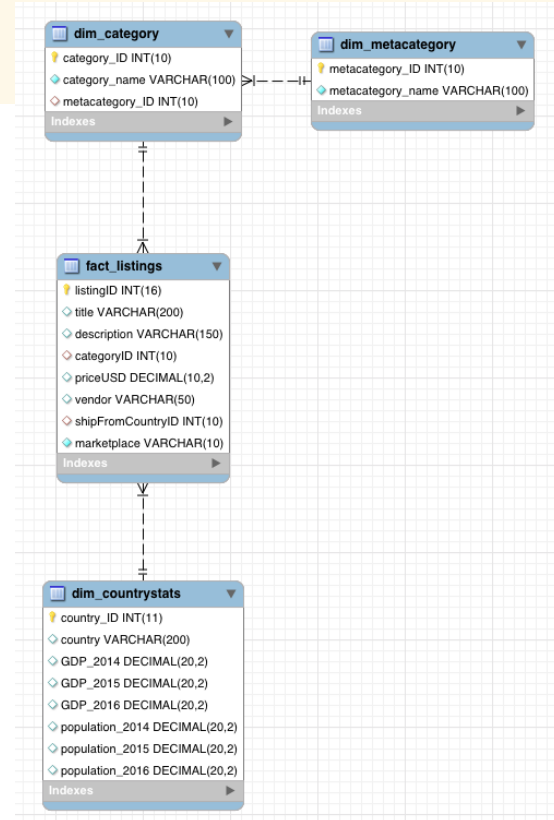


The screenshot shows a database query interface. At the top, a SQL query is entered: `SELECT * FROM dim_countrystats;`. Below the query, a toolbar includes options for 'Filter Rows', 'Search', 'Edit', and 'Export/Import'. The main area displays a 'Result Grid' with 8 columns: country_ID, country, GDP_2014, GDP_2015, GDP_2016, population_2014, population_2015, and population_2016. The grid contains 15 rows of data, starting with Aruba and ending with Burundi.

country_ID	country	GDP_2014	GDP_2015	GDP_2016	population_2014	population_2015	population_2016
1	Aruba	0.00	0.00	0.00	103795.00	104341.00	104822.00
2	Afghanistan	20616104298.00	19215562179.00	19469022208.00	32758020.00	33736494.00	34656032.00
3	Angola	126777000000.00	102962000000.00	95335111741.00	26920466.00	27859305.00	28813463.00
4	Albania	13228244357.00	11335264967.00	11863865978.00	2889104.00	2880703.00	2876101.00
5	Andorra	3350736367.00	2811489409.00	2858517699.00	79223.00	78014.00	77281.00
6	Arab World	290662000000.00	256330000000.00	250470000000.00	390043028.00	398304960.00	406452690.00
7	United Arab Emirates	403198000000.00	357949000000.00	348743000000.00	9070867.00	9154302.00	9269612.00
8	Argentina	52632000000.00	58471100000.00	54547600000.00	42981515.00	43417765.00	43847430.00
9	Armenia	11609512940.00	10553337673.00	10572298342.00	2906220.00	2916950.00	2924816.00
10	American Samoa	643000000.00	659000000.00	658000000.00	55437.00	55537.00	55599.00
11	Antigua and Barbuda	1280133333.00	1364863037.00	1460144704.00	98875.00	99923.00	100963.00
12	Australia	145960000000.00	134538000000.00	120462000000.00	23504138.00	23850784.00	24210809.00
13	Austria	44188500000.00	38206600000.00	39080000000.00	8541575.00	8633169.00	8731471.00
14	Azerbaijan	75244166773.00	53074370486.00	37847715736.00	9535079.00	9649341.00	9757812.00
15	Burundi	3093647227.00	3066681387.00	3007029030.00	9891790.00	10199270.00	10524117.00

Dimensional Model

A snowflake schema was utilized for the data warehouse because it allows for faster analysis and querying as compared to a normalized, OLTP-style data structure

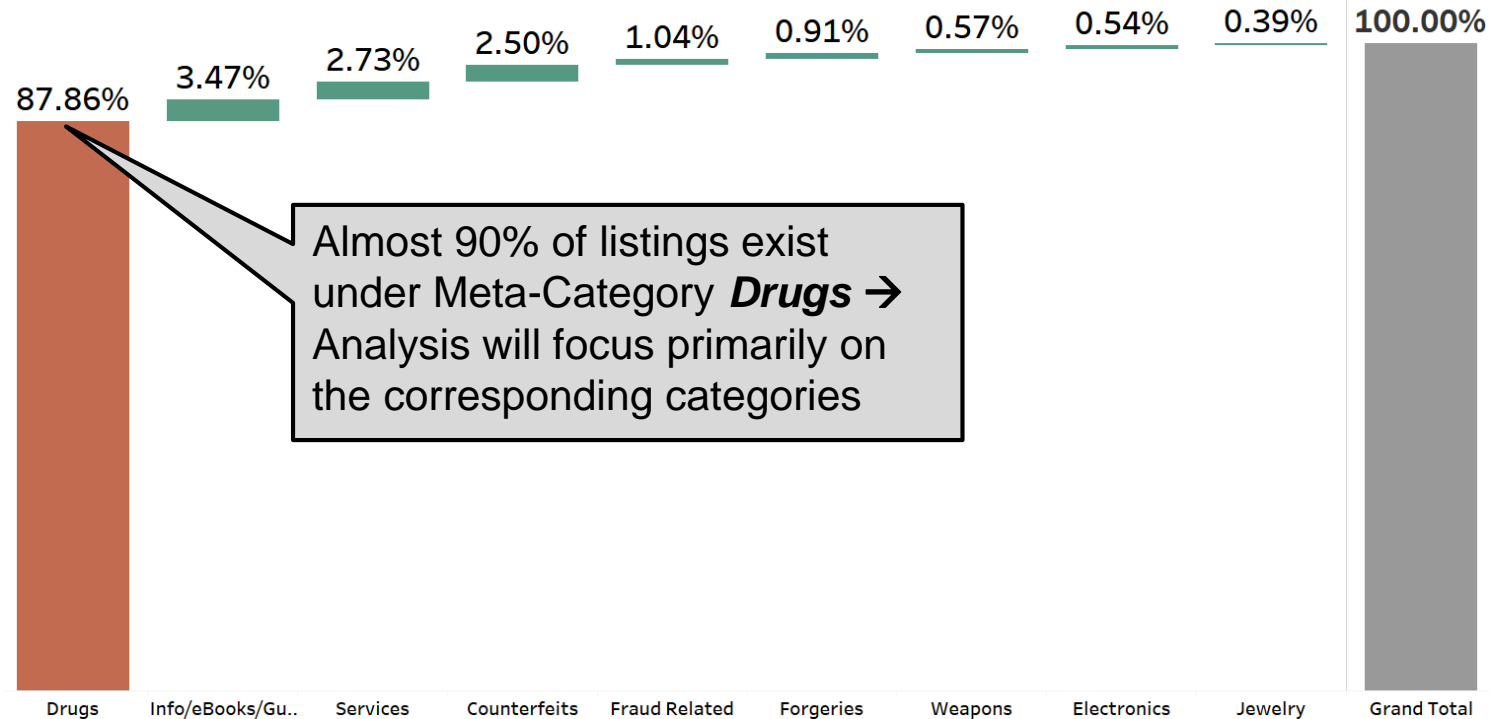


Findings – Disclaimer

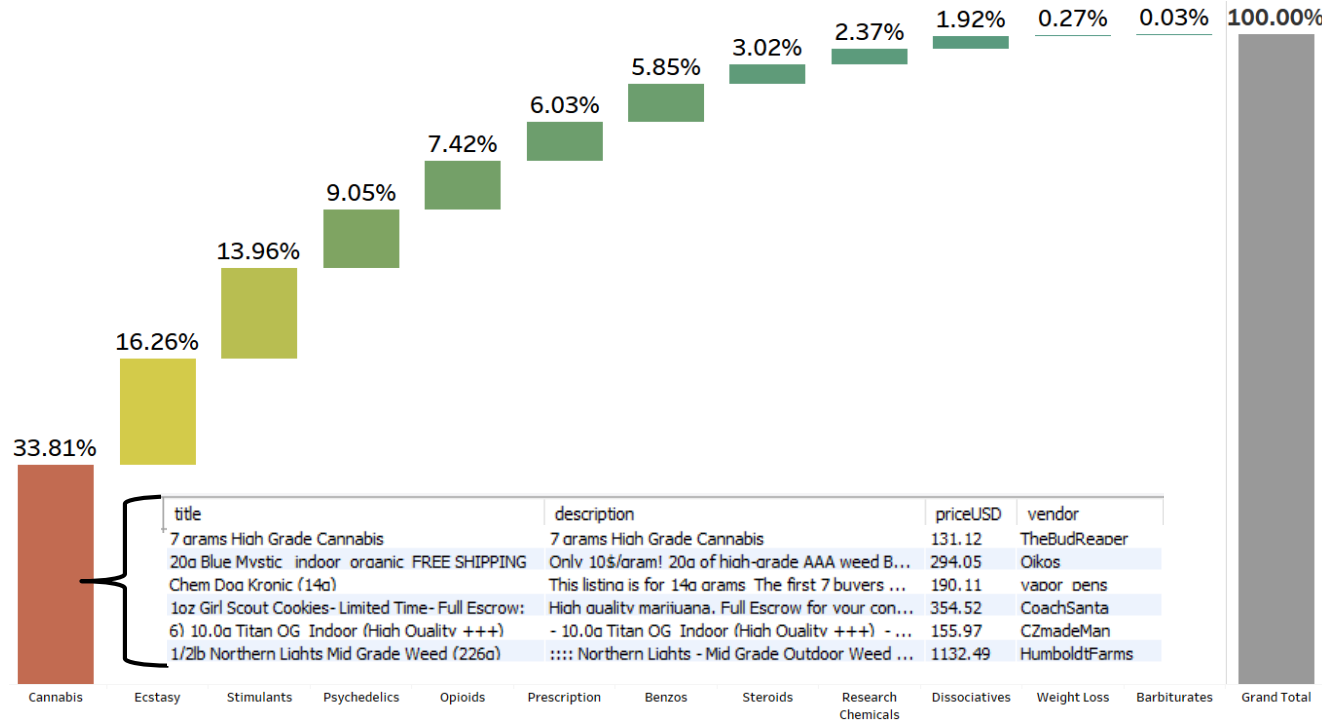
- All analysis that follows is only representative of the data used in this project and is not guaranteed to be generalizable across new darknet marketplace data
- Please consider ETL steps outlined in **Methodology** section when interpreting results of analysis



Meta-Category Breakdown*



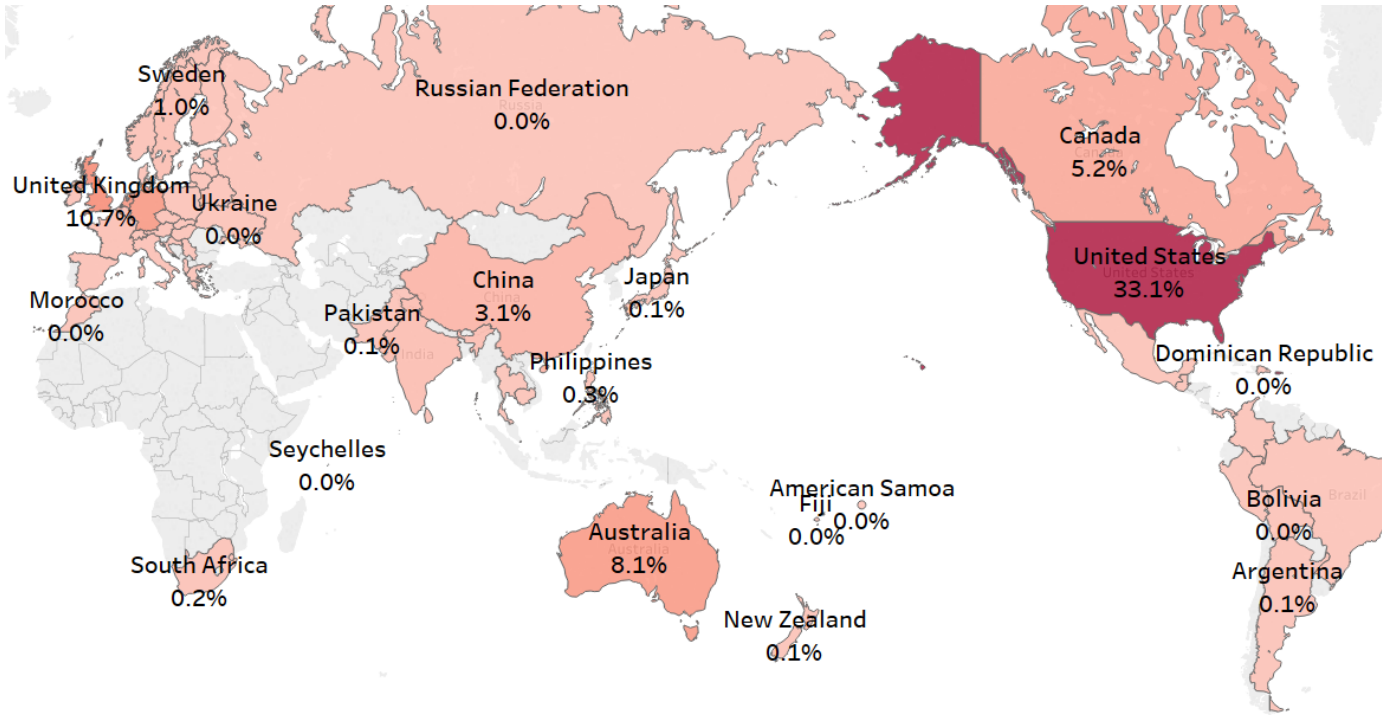
Drug Categories*



Key Points

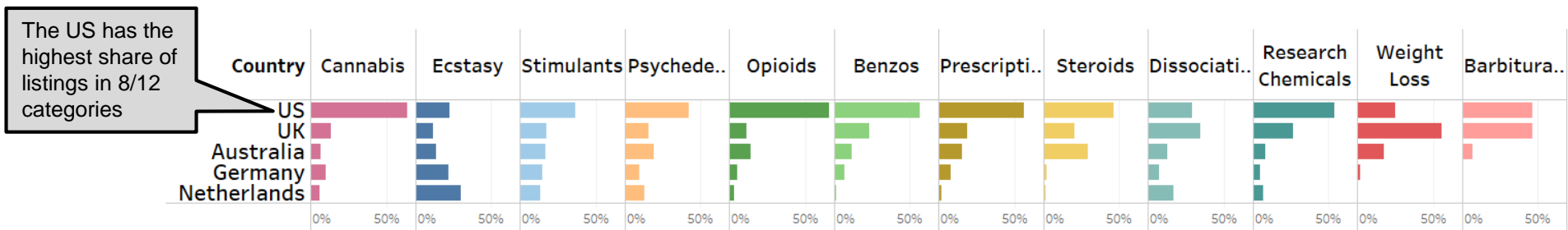
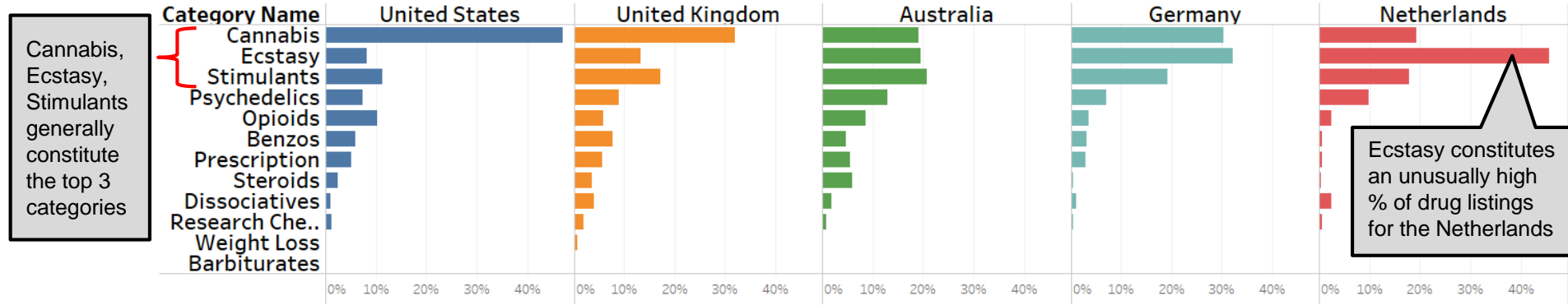
- Listing descriptions are varied and denote different item quantities; prices exhibit a large range
- Top 5 categories account for ~80% of classified Drug listings
- Do products listed in any top category originate from a few primary countries?

Percentage of Drug Listings, By Country

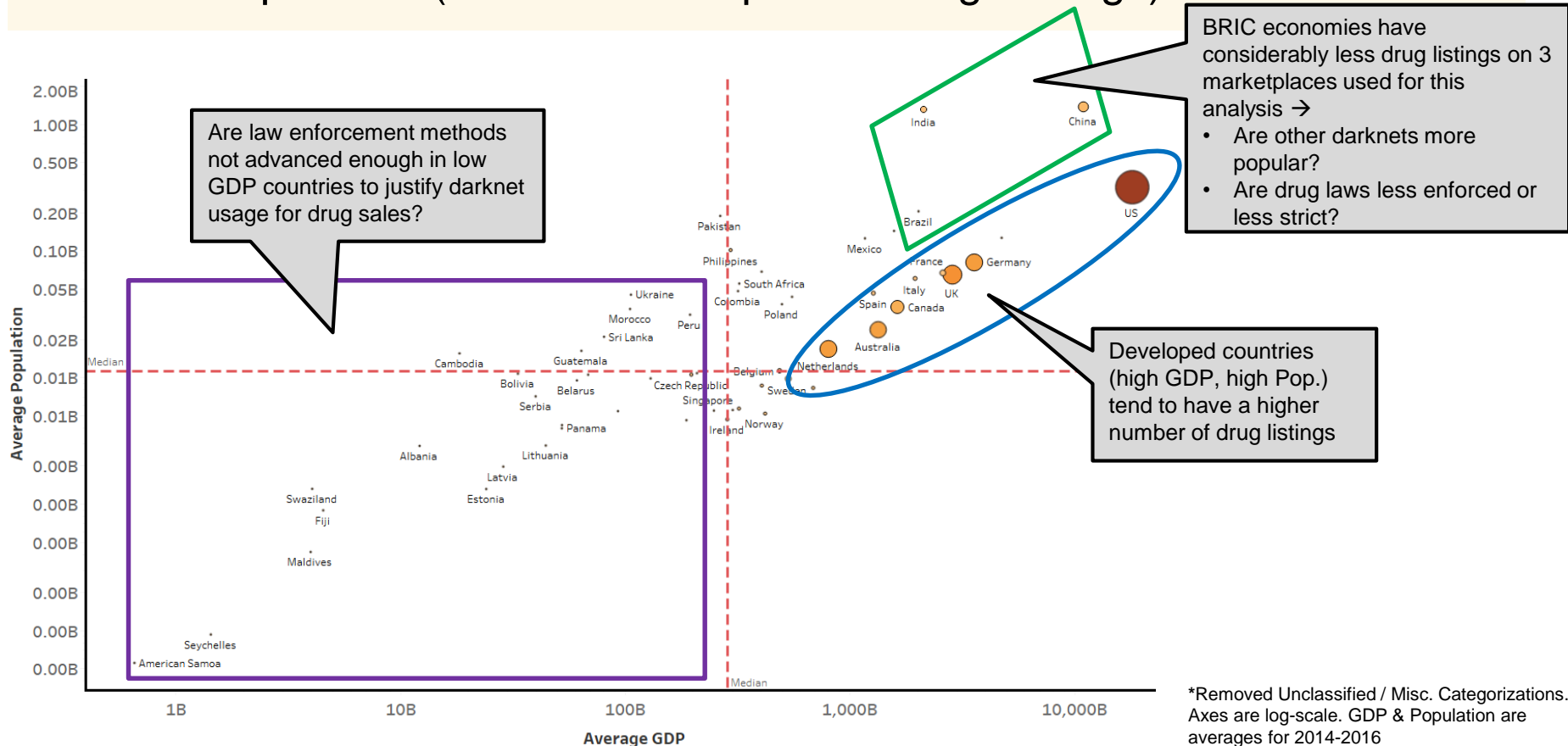


Only 5 countries (U.S., Australia, U.K., Germany, Netherlands) seem to be listed as an origin point for drugs a sizeable amount of the time

Top 5 Countries – Drill Down*



GDP vs. Population (Bubble Size equals # Drug Listings)



Recommendations

- Current database allowed for exploratory analysis
 - Data from more darknet marketplaces can determine if observations made in this analysis are generalizable
- Data was not consistent across different darknet marketplaces
 - Similar time frames, product UOM broken out, etc. will yield more robust comparisons
 - Standardized data will enable predictive analytics
 - Regression – predict product prices based on factors such as item description length, vendor rating, product category
 - Classification – Using existing database of listings, predict where new listings will originate from (country, vendor)
- Listings data can only tell so much of the full story - having access to actual transactions data facilitates much more sophisticated analyses



Lessons Learned

- A good project plan keeps things on track and moving along
- Thinking of the business use case is helpful when trying to combine multiple data sets that are not in the same format
- ETL *does* take the majority of time!
- Ensure your database has the appropriate level of security...



Appendix – Data Sources, 1 of 2

- Agora Listings: <https://www.kaggle.com/philipjames11/dark-net-marketplace-drug-data-agora-20142015>
- Hansa & Valhalla Listings: <https://polecat.mascherari.press/onionscan/dark-web-data-dumps>
- Country GDP: <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2015&start=2014>
- Country Population: https://data.worldbank.org/indicator/SP.POP.TOTL?end=2016&start=2006&year_low_desc=false



Appendix – Data Sources, 2 of 2

Agora

Vendor	Category	Item	Item Description	Price	Origin	Destination	Rating	Remarks
EmeraldTriangle	Drugs/Cannabis/Weed	1/8 oz Green Ribbon - FREE SHIPPING	**** READ OUR PROFILE BEFORE ORDERING **** FAILURE TO 0.14805414 USA		USA ONLY		4.88/5	
joeybagadonuts	Drugs/Cannabis/Weed	1 oz (29 plus grams) BLACK TRAINWRECK -	*ALL ORDERS MUST FE we hate to require this but have to -1 (29.25 1.34007154 USA		USA		5.000/5	
bigfootbuds	Drugs/Cannabis/Weed	1/2 lb Yeti's Dream	Top Shelf Yeti's Dream (loved from Coast to Coast). Our own special 8.80761594 USA				5.00/5	Average price
BossCalBud	Drugs/Cannabis/Weed	1 Oz(28g) - Vanilla Kush - FE DISCOUNT - 2-	I have 2 Ounces of Vanilla Kush. This stuff is CRYSTALLY and TASTY. A 0.87974904 USA		USA		5.00/5	
blow	Drugs/Cannabis/Weed	H.P. (227g) AK47 - A+	Strain: AK47 Type: Hybrid Sativa-dominant THC %: 16-22% Grown: In 4.82154671 Canada		Worldwide		4.83/5	Average price
StrattonOakmont	Drugs/Cannabis/Weed	56 Grams Cheese Mid Shelf Premium Marij	56 Grams of StrattonOakmont's Top Shelf Washington Medical Marij 1.37187684 USA		USA		4.987/5	
AngelEyes	Drugs/Cannabis/Weed	1/4 Ounce The Big Lebowski	The Big Lebowski (2-3 week cure) Indica/sativa blend. Over the year 0.28515374 USA		USA		4.99/5	
Utopic	Drugs/Cannabis/Weed	56g 2oz [Utopic's] TRUSTED AAA Indoor	[INSANE STEALTH GUARANTEE EVERY ORDER] Properly Purged Wax 1.85076294 USA		USA		4.846/5	
Utopic	Drugs/Cannabis/Weed	3.5g [Utopic's] TRUSTED Grade A Indoor 'V	[INSANE STEALTH GUARANTEE EVERY ORDER] Properly Purged Wax 0.12395404 USA		USA		4.846/5	
kraeutergarten	Drugs/Cannabis/Weed	Sgr Peace Maker TOP Quality Weed	kraeutergarten präsentierte: Peace Maker TOP Qualität Sort: Ind 0.2750335; Germany		Europe		4.99/5	
GreenSurfer	Drugs/Cannabis/Weed	Power Plant 14G	Power Plant comes from Dutch Passion Seed Company and is derived 0.7485232; UK		UK		4.99/5	
hanspoel	Drugs/Cannabis/Weed	25g Schnittraste von PP weed	25g Schnittraste Endlich wieder da Schnittraste!!! (meiste ist Reste 0.3261361; Germany		Germany		4.99/5	

Hansa

Timestamp (when the listing was scrapped)	Marketplace ID	Title	Vendor	Price	Marketplace Category	Ships from Country
2016-12-11 20:13:54.654635085 -0500 EST	"18365"	"50gr. FUB-AKB"	"Alaurizen"	"USD 340.00"	"Drugs"	"China"
2016-12-11 15:34:46.900547093 -0500 EST	"12228"	"Get a FREE iPhone 6s III"	"smart666tiger"	"USD 11.99"	"Electronics"	
2016-12-11 18:19:59.667791589 -0500 EST	"37713"	"The Basics of Information Security 2nd edition 2014"	"pckabml"	"USD 2.99"	"Guides & Tutorials"	
2016-12-11 19:00:59.930441692 -0500 EST	"9463"	"Oxazepam 50x (30MG) Pills"	"Discover"	"USD 115.00"	"Drugs"	"United States"
2016-12-11 20:06:18.05821968 -0500 EST	"30068"	"10g - AH Special Snow White Cocaine - 90% Pure"	"Coolblue"	"USD 517.56"	"Drugs"	"Europe (EU)"
2016-12-11 16:27:33.76323067 -0500 EST	"47286"	"Ruile - Day Date 40 Yellow Gold 228238 Black Diagonal M	"RepaAA"	"USD 379.00"	"Counterfeits"	"Hong Kong"
2016-12-11 17:14:22.315626949 -0500 EST	"19454"	"50 Gram Dutch Quality Ketamine S-Ketamine"	"DutchDrugs"	"USD 1	"Drugs"	"Worldwide"
2016-12-11 17:54:36.324916829 -0500 EST	"33031"	"10g of High Quality #3 Heroin"	"sargon"	"USD 628.77"	"Drugs"	"Worldwide"
2016-12-11 18:50:58.325840342 -0500 EST	"38234"	"PremiumGfs.com - [LIFETIME PORN PREMIUM ACCOUNT]"	"VideoK"	"USD 4.99"	"Digital Goods"	
2016-12-11 18:48:51.425465524 -0500 EST	"55008"	"20 gr super heroin from pakistan AAA++"	"goldendrugs"	"USD 792.18"	"Drugs"	"Netherlands"
2016-12-11 15:48:02.684193851 -0500 EST	"10677"	"Connecticut - CT SSN + DOB FULLZ"	"Zloy3"	"USD 1.25"	"Fraud Related"	
2016-12-11 16:11:46.962031811 -0500 EST	"45065"	"100x MicroDots - 150ug JOR #4 LSD"	"BrainCandy"	"USD 289.00"	"Drugs"	"United States"
2016-12-11 18:39:45.281296982 -0500 EST	"54392"	"60x Bupropion Extended-Release (SR) 150 mg"	"ThisIsTheWhy"	"USD 10.20"	"Drugs"	"United States"

Valhalla

Marketplace Title	Vendor	Price	Ship From Counter
10086 Hacking for Profit: Credit Card Fraud A Beginners G	junkiepig666	2.73 EUR	Germany
10087 Introduction to Social Engineering	junkiepig666	2.73 EUR	Germany
10096 MST Morphine sulphate tablets 100 mg =30\$	pure12	29.14 EUR	United Kingdom
10097 MORPHINE SULPHATE TABLETS (CONTINUS) 30\$ =	pure12	29.14 EUR	United Kingdom
10099 Learn The Basics of Ethical Hacking and Penetration	junkiepig666	18.22 EUR	Germany
10100 How I Sell \$7500 Month Online Without Google or	junkiepig666	18.22 EUR	Germany
10102 MORPHINE SULPHATE TABLETS CONTINUS 30mg x	pure12	29.14 EUR	United Kingdom
10103 MST CONTINUS branded product ,10 X 100 mg tab	pure12	174.85 EUR	United Kingdom
10116 SUBUTEX BUPRENORPHINE 3 X 8mg = 55\$ branded	pure12	51.91 EUR	United Kingdom
10202 Udemy -Advanced White Hat Hacking & Penel	junkiepig666	18.22 EUR	Germany

GDP

Country Name	2014	2015	2016
Aruba			
Afghanistan	2.0616E+10	1.9216E+10	1.9469E+10
Angola	1.27E+11	1.03E+11	9.5335E+10
Albania	1.3228E+10	1.1335E+10	1.1864E+10
Andorra	3350736367	2811489409	2858517699
Arab World	2.91E+12	2.56E+12	2.50E+12
United Arab Emirates	4.03E+11	3.58E+11	3.49E+11
Argentina	5.26E+11	5.85E+11	5.45E+11
Armenia	1.161E+10	1.0553E+10	1.0572E+10
American Samoa	643000000	659000000	658000000
Antigua and Barbuda	1280133333	1364863037	1460144704
Australia	1.46E+12	1.35E+12	1.20E+12
Austria	4.42E+11	3.82E+11	3.91E+11
Azerbaijan	7.5244E+10	5.3074E+10	3.7848E+10

Population

Country Name	2014	2015	2016
Aruba	103795	104341	104822
Afghanistan	32758020	33736494	34656032
Angola	26920466	27859305	28813463
Albania	2889104	2880703	2876101
Andorra	79223	78014	77281
Arab World	390043028	398304960	406452690
United Arab Emirates	9070867	9154302	9269612
Argentina	42981515	43417765	43847430
Armenia	2906220	2916950	2924816
American Samoa	55437	55537	55599
Antigua and Barbuda	98875	99923	100963

Findings

Recommendations

Lessons Learned

Appendix