**Objective of this document is to provide context about variations program – background, impact & current status, and detail out high level requirements**

1. **Background**

Customers use variations to find new products, reorder known products, and explore different product formats. Vendors/sellers use variations to illustrate selection, define brand, encourage product discovery, and drive sales. Confusing CX, poor inbound data quality, degrading data integrity, minimal governance, and inadequate information architecture limit the current experience by requiring a high cognitive load to process seemingly simple tasks. Variation families consists of near identical products varying by specific item attributes (called themes). Most popular themes are size and color for shoes, volume and fragrance for detergents, size and flavor for groceries. Products may come in different sizes, flavors, scents, colors, etc. and complexities arise when products vary across many dimensions, have numerous size/pack options, or have many variants per dimension.

Below mentioned are the primary sources impacting detail page variations –

1) Inconsistent families: Unrelated items can be grouped in the same variation family due to bad data or variation relationship. A seller/vendor can knowingly or unknowingly create an illegitimate parent-child relationship between the parent of a popular variation family and an unrelated ASIN by adding a SKU on both thus forcing IOP-Twistering of the family. As a result, the unrelated children benefit from the reviews of the original variation family [Refer to Appendix C – example use case 7]

2) Un-normalized attributes: Un-normalized variation theme attribute values create near duplicate or inconsistent values in detail page-twisters (Refer to Appendix X). We refer to such families as un-normalized families. The root cause of un-normalized data is lack of quality control at data entry and lack of normalization at later stages. [Refer to Appendix C – example use case 1]

*Quoting a customer "There's not a per ounce price on every single one of them. Some of them are different sizes. And they're mixed up whether they are in cases of 6 or a 1 individual. This would take me a while. I'd get out my calculator and fuss over this quite a bit. Or I'd go to Super Supplements because I think they're cheaper." – Kay (heavy Amazon shopper) abandons the task of shopping for a specific brand of toothpaste.*

1. **Business Impact**

In a preliminary study of 650 U.S. Consumables (Baby, Beauty, and HPC) ASINs, 80% had variations and of those 48% had poor quality variations. The non-error sample had 7% higher shipped units/GV (average) suggesting ~$1.2B lost annual GMS (Appendix A). A subsequent pilot for Consumables Coffee category, revealed a conversion improvement by 7.7% at a family level for the ASINs with resolved errors as compared to the control set (Appendix B). We identified Coffee category for the Pilot in order to test the Find and Fix solution. Based on the sizing of these defect types revealed by our study, we prioritized size related fixes displayed under the size dimension of the detail page twister. During the beginning of the pilot (wk. 38), coffee category had 35.8% size related variation defects. We were able to setup scalable mechanisms to identify and monitor the Coffee category ASINs for 6 size defects. We leveraged the work done by the New Item Setup (NIS) teams within RBS and defined the process to fix the identified defects. By the end of the pilot (wk. 46) we were able to reduce the defect rate from 35.8% to 7.4%. During the process we also experienced continuous ingress where the cleaned up families were re-infected suggesting that cleanup is a continuous process and will require continuous fixing of ASINs. Based on the success of the coffee pilot, we have setup the operations to reduce variation defects for US CL (refer goals for the program). As of 2nd week of Feb, we on boarded 40 categories (across drugstore and grocery) for defect fixing, thereby expanding our coverage to 87.3% of scoped ASINs. We reduced the GMS Weighted defect rate down from 40.5% in week 1 to 26.4% by end of Week 6’ 2019 [Refer to Appendix D for Variations Dashboard].

For Softlines, an audit on a random sample of 900 ASINs, revealed 10.9% and 31.1% defective ASINs for Apparel and Shoes respectively, suggesting an annualized impact to OPS of $336M for US Shoes and Apparel alone. With over 289M buyable ASINs (vs. 21M in Hardlines), 46.9% of which are less than one year old (vs. 10.4% in non-Softlines product families), and 80.8% are variations (vs. 28.8% in non-Softlines), Softlines has one of the largest and most variated catalogs across Amazon. Despite the importance of variation relationships, management of variation quality and performance has been ad-hoc and escalation based. [Refer to Appendix – A for details on financial impact calculations]

1. **Teams & Tools**

Variated ASINs touch a variety of systems, teams, and tools throughout their lifecycle. Data quality can be impacted at various places throughout an ASIN lifecycle including: Ingestion, ASIN Creation, Variations Relationship Creation, Catalog Update, and Website Display stages. Variation UX can be impacted by changes in search, detail page, and reviews on both desktop and mobile platforms. To address this problem holistically, RBS and U.S. Consumables RO&I have partnered to co-own development and delivery of this Variations program. However, we have identified a number of additional stakeholders that will own key themes. Additional identified program stakeholders include: ASCS (IoP, Catalog Programs & Governance (CPG) /AVS), Selection Central Platform (SCP) team, Marketplace Listing Metadata Service (MLMS). Secondly, we have identified a number of operational stakeholders and teams running key systems/tools who would be impacted by changes to variations. Operations stakeholders include: NIS, CMT, TAM, VCM, CX Ops, and Retail. System/tool stakeholders include: AVS, NIS, CMT, Variations Wizard, SPOT, Darwin, ASIN Toughening, ACT, Selection Central, RBS Sherlock, and RBS Auto-Enrich Tool.

1. **2019 Goals for Variations program**

* Reduce detail page variation size and PPU catalog defects from XX% to <5% weighted by GMS for Band A-B NA Consumables by 03/31/2019
* Reduce detail page variation size and PPU catalog defects from XX% to <5% weighted by GMS for Band A-B across all Eng. speaking MPs for Consumables and Softlines by 10/31

# Variations – Use Cases

Below are the list of variation use cases that are in roadmap for the variations program [Refer Appendix I - for use case prioritization methodology)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No** | **Variation Defect type** | **Priority** | **Catalog Attribute** | **Description** |
| 1 | Duplicates in size | 1 | Size | Size variation error where multiple values of the same size are present in the twister |
| 2 | Inconsistent unit of measure in size | 1 | Size | Size variation error where values have inconsistent units within the same twister |
| 3 | No unit of measurement in size | 1 | Size | Size variation error where value does not have a unit of measurement |
| 4 | Incorrect information in size | 1 | Size | Size variation error where non-size variation information is included in size attribute |
| 5 | Missing PPU | 2 | Unit Count,  Unit Count type | PPU error where PPU attribute is missing in twister |
| 6 | Incorrect unit of measure in PPU | 2 | Unit Count type | PPU error where unit of measure is incorrect for the product group |
| 7 | Inconsistent unit of measure in PPU | 2 | Unit Count type | PPU error where unit of measure is inconsistent within the variation family |
| 8 | Incorrect information in color | 3 | Color | Color attribute populated with style, size, scent, flavor and other irrelevant values |
| 9 | Pack values in size | 4 | Size | Size variation error where value is a pack size and not a size |
| 10 | Pack values in other variating attributes | 4 | Flavor, Style, Color, Scent | Pack values in other variating attributes flavor, scent, style, color |
| 11 | Incorrect Variation theme | 4 | Variation  Dimension Order | Theme of parent/variation family is incorrect |
| 12 | Incorrect information in other variating attributes | 5 | Flavor, Style, Scent | Flavor, scent, style attributes are populated with irrelevant values |

# KPIs to measure success

|  |  |
| --- | --- |
| **System/Program** | **Metrics** |
| **Overall Variations Program** | **Defect Metrics:**  GMS Weighted Defect Rate  Ingress % - # of Ingress ASINs when compared to clean catalog from previous week  Respawn % - # of Respawn ASINs when compared to clean catalog from previous week  **Business Metrics:**  Conversion (Shipped units/GV)  Incremental OPS |
| **Sherlock** | Total number of ASINs Reviewed/Audited  Total Actionable defects identified (AER)  Recall % |
| **ACE** | Correction accuracy % (# ASINs with correct values/# ASINs audited)  Correction coverage % (# ASINs corrected by ACE/# Input ASINs for ACE) |

1. **CAPABILITIES/EPICS**

As we move forward with variations program, we would require scalable mechanisms to solve this problem for more (MPs, GLs, GV BANDs), with less human intervention and better customer experience. We need capabilities for Variations program that allows us to prioritize ASINs rationally, identify defects systematically at scale, correct defects through automated approach, and provide mechanisms for humans to validate/correct ASINs that are not corrected through the system. These epics are numbered in the sequence of variations process flow. These are independent capabilities that could be part of any of the RBS BTO Stack (Sherlock/ACE/ZEUS) or any other amazon platforms. By nature, some tasks will be interdependent on some others but those technical dependencies will have to be worked out closer to development of these user stories.

1. Variation ASIN input – We need to build approach to retrieve buyable variation families (both Child & Parent ASINs) for either an identified input criteria (e.g. BAND A-B ASINs) (or) for a list of uploaded ASINs in specific format
2. Defect Auto Detection - Defect detection for above listed variation use cases can happen through deterministic rules or through ML based approach. For defect detection through rules, we will need to build a comprehensive and scalable approach for valid values/inclusion list for each of the variating attributes (Size/Flavor/Color/Scent/Style). Through the variation US CL program, we have built ideal size values at ITK level and preferred variation themes at sub category level for 679 sub categories (Appendix E). We need to build similar mapping tables for other variating attributes.
3. Defect validation – We need to build capability to manually validate system identified defects. Currently, we have mechanism in Sherlock to download reports and validate them through excel sheets. This is not be a sustainable approach going forward. We would need to build a user interface to validate defects efficiently and collect data systematically in the BTO world
4. Defect auto correction can have four sub epics
   1. Input Feed – We need to build an approach to pick up tasks from BTO Bus (or) onboard tasks for a list of uploaded ASINs in specific format
   2. Data sourcing – For Incorrect values or no value in variating attributes, we need to build an approach to source normalized data from product image/title/description and for un-normalized values the system should refer normalization table to source normalized values. Normalization table (Appendix G) has been built during US CL pilot through a manual mapping process. We need to build a scalable approach to find normalization values for all categories across the catalog
   3. Data Validation - We need to build capability to sideline ASINs and manually audit the system generated correction recommendations
   4. Processing – We need to build capability to update validated attribute values into production catalog
5. Defect manual correction – We need to build capability to optimally allocate sidelined ASINs for RBS/m-Turk correction process or selling partner correction process
6. Metrics & Monitoring –We need to have required reporting mechanisms for users to view required metrics. List of metrics needed and mock reports are attached in Appendix X.
7. **User Stories**

|  |  |  |  |
| --- | --- | --- | --- |
| **Epics** | **System** | **User Stories** | **Releases** |
| Variations ASIN input (As user X, I want ASINs frequently seen by customers to have consistent and good quality variations) | SHERLOCK | As user X, I want the system to retrieve buyable variation families (both Child & Parent ASINs) for the identified input criteria (BAND A-B ASINs and their variation families, so that I can send the list of ASINs for auto detection process. | TBD |
|  | SHERLOCK | As user X, I want the system to provide mechanism to manually upload list of ASINs, so that I can send the list of ASINs for auto detection process during adhoc situations | TBD |
| Defect Checklist | SHERLOCK | As user X, I want the system to retrieve all relevant attribute values, so that detection checks can be applied on them | TBD |
| Defect auto detection | SHERLOCK | As user X, I want the system to apply all relevant rules or ML modules, so that variation defects can be identified for the input ASINs | TBD |
|  | SHERLOCK | As user X, I want the system to capture unique record for each Child ASIN/defect (or) for each PARENT ASIN/defect, so that I can send the defects downstream for manual validation | TBD |
|  | SHERLOCK | As user X, I want the system to provide an UI with information on defects identified, so that I can manually validate the defects | TBD |
| Defect Validation | SHERLOCK | As user X, I want the system to create an auto correction task on the BTO BUS for all manually validated defects, so that I can send the defects downstream for correction | TBD |
| Defect auto correction | ACE | As user X, I want the system to onboard variation correction task and pick up defects from BTO BUS or onboard tasks through an excel file upload | TBD |
|  | ACE | As user X, I want the system to source attribute values from image/description of the product for ASINs with no attribute values or incorrect attribute values, so that I can use the correct attribute values for backfilling | TBD |
|  | ACE | As user X, I want the system to normalize attribute values and make it consistent for all variation family ASINs so that I can use the consistent and normalized attribute values for backfilling | TBD |
|  | ACE | As user X, I want the system to provide information required for a human to perform Manual Correction when it fails to Auto Correct or where we want to perform an audit of an ASIN before correction | TBD |
|  | ACE | As user X, I want the system to publish the validated attribute values in the catalog, so that customer is able to see the correct and consistent values in detail page twister | TBD |
|  | ACE | As user X, I want the system to create manual correction task in BTO BUS for ASINs failed to correct, so that I can allocate the same for manual corrections | TBD |
| Defect manual corrections | ZEUS | As user X, I want the system to pick up the tasks from BTO BUS, prioritize and allocate tasks for associates/m-Turk based on defined criteria, so that pending defects could be manually corrected | TBD |
|  | ZEUS | As user X, I want the system to allocate relevant tasks to selling partners, so that pending defects are corrected |  |
| Metrics & Monitoring  (Need to brainstorm with platform PMs to arrive at the right metrics) | SHERLOCK | Tracking and reporting following metrics for Sherlock:  #Input ASINs, # ASINs audited by Sherlock, # of ASINs sidelined by Sherlock, # ASINs manually validated, #defective ASINs, Defect rate %, GMS weighted defect %, Defect types (# of Defects), # Tasks passed on to ACE for corrections | TBD |
|  | ACE | Tracking and reporting following metrics for ACE:  # Tasks received by ACE for corrections, # tasks for which attributes sourced, # tasks for which attributes could not be sourced, # tasks corrected by ACE, # tasks sidelined for manual correction | TBD |
|  | ZEUS | Tracking and reporting following metrics for ACE:  # tasks received by ZEUS for manual corrections, # tasks allocated for manual corrections, # tasks allocated to selling partners, # of manual corrections completed, # corrections pending | TBD |

# Functional Requirements for Sherlock (Defect Identification)

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| **F1** |  | **Duplicates in Size** |
| F1.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F1.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or without a size variation dimension |
| F1.3 |  | The System must retrieve size attribute details for child ASINs |
| F1.4 |  | The System must identify all un-normalized unit of measure values causing duplicates in twister as defect by validating with normalization table (Appendix E) and mark defect type as “Duplicates in Size – UOM” (e.g. Ounce, Oz, Ounces within the same variation family) |
| F1.5 |  | The system must identify conversion size value duplicates in twister as defect and mark defect type as “Duplicates in Size – Conversion” (e.g. 500 g, 0.5 kg, 0.5 kgs within the same variation family) |
| F1.6 |  | The system must identify un-normalized numeric values causing duplicates in twister and mark defect type as “Duplicates in Size – unit quantity” (10 Ounce, 10-Ounce, 10.Ounce) |
| F1.7 |  | The System must mark defect at child level |
| **F2** |  | **Inconsistent unit of measurement in Size** |
| F2.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F2.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or without a size variation dimension |
| F2.3 |  | The System must retrieve size attribute details for child ASINs |
| F2.4 |  | The System must identify inconsistency in unit of measure within the variation family as defect (e.g. variation family with Ounce, Pound, Grams as unit of measure within the same family) |
| F2.5 |  | The System must identify inconsistency in unit of measure within the family when unit of measure is un-normalized as defect (e.g. variation family with Ounce, lbs., gram as unit of measure within the same family) |
| F2.6 |  | The System must not mark un-normalized unit of measures of same measurement as defect (e.g. variation family with Ounce, Oz, Oz. as unit of measures) |
| F2.7 |  | The System must mark defect type at parent level |
| **F3** |  | **No unit of measurement in Size** |
| F3.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F3.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or without a size variation dimension |
| F3.3 |  | The System must retrieve size attribute details for child ASINs |
| F3.4 |  | The System must identify only numeric and decimal values in size attribute as defect |
| F3.5 |  | The System must mark defect type at child level. The defect type should be named as “No unit of measurement in size ” |
| **F4** |  | **Incorrect information is Size** |
| F4.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F4.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or without a size variation dimension |
| F4.3 |  | The System must retrieve size attribute details for child ASINs |
| F4.4 |  | The System must identify ASINs with size values outside the standard size mapping as defect |
| F4.4.1 |  | The System must check child ASIN size value with ITK level ideal size mapping value to validate (Appendix G) |
| F4.5 |  | The System must mark defect type at child level. The defect type should be named as “Incorrect information in Size” |
| **F5** |  | **Pack Values in Size** |
| F5.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F5.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or without a size variation dimension |
| F5.3 |  | The System must retrieve size attribute details for child ASINs |
| F5.4 |  | The System must identify ASINs with “Pack & other related strings” in size attribute as defect (Appendix H) |
| F5.5 |  | The System must mark defect type at child level. The defect type should be named as “Pack in Size” |
| **F6** |  | **Pack Values in other variating attributes** |
| F6.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F6.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore |
| F6.3 |  | The System must retrieve flavor/color/style/scent attribute details for child ASINs based on the variation dimension order of the family |
| F6.4 |  | The System must identify ASINs with “Pack & other related strings” in relevant flavor/color/style/scent attribute as defect |
| F6.5 |  | The System must mark defect type at child level. The defect type should be named as “Pack in other variating attributes” |
| **F7** |  | **Incorrect Variation theme** |
| F7.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F7.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore |
| F7.3 |  | The System must retrieve variation dimension order attribute details for Parent ASINs |
| F7.4 |  | The System must identify Parent ASINs with variation themes outside the variation theme mapping table as defect |
| F7.5 |  | The System must mark defect type at Parent level. The defect type should be named as “ Incorrect variation theme” |
| **F8** |  | **Missing unit count & unit count type in PPU** |
| F8.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F8.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore |
| F8.3 |  | The System must retrieve unit count & unit count type attribute details for child ASINs |
| F8.4 |  | The System must identify ASINs with unit count AND/OR unit count type as NULL as defect |
| F8.5 |  | The System must mark defect type at Child level. The defect type should be named as “ Missing PPU” |
| **F9** |  | **Incorrect unit of measure in PPU** |
| F9.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F9.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore |
| F9.3 |  | The System must retrieve unit count type attribute details for child ASINs |
| F9.4 |  | The System must identify ASINs with unit count type value outside size mapping table as defect |
| F9.5 |  | The System must mark defect type at Child level. The defect type should be named as “ Incorrect unit of measure in PPU” |
| **F10** |  | **Inconsistent unit of measure in PPU** |
| F10.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F10.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore |
| F10.3 |  | The System must retrieve unit count type attribute details for child ASINs |
| F10.4 |  | The System must check for consistency of unit measure across family, if found inconsistent mark as defect |
| F10.5 |  | The System must mark defect type at Parent level. The defect type should be named as “ Inconsistent unit of measure in PPU” |
| **F11** |  | **Incorrect information in Color** |
| F11.1 |  | The System shall pull child ASINs & variation parent for the input criteria defined |
| F11.2 |  | The System shall abandon ASINs which are not active and/or non-buyable and/or not part of variation family anymore and/or not having color variation dimension order |
| F11.3 |  | The System must retrieve color attribute details for child ASINs |
| F11.4 |  | The System must identify ASINs with color values outside the standard color table mapping as defect |
| F11.5 |  | The System must mark defect type at child level. The defect type should be named as “ Inconsistent information in Color” |
| **F12** |  | **Incorrect information in Other variating attributes** |
|  |  | - Solution not defined yet |

# Success criteria for Sherlock

Sherlock should identify variation defects with a precision (AER) of >80% and a recall of >90%.

# Functional requirements for ACE (Defect Corrections)

\*\* Refer Appendix J for detailed SOP for variation corrections

\*\* Two functional use cases have been documented below, based on feedback and discussion rest of the functional requirements will be updated by XX/XX

|  |  |  |
| --- | --- | --- |
| **ID** | **Priority** | **Description** |
| **F1** |  | **Duplicates in Size** |
| F1.1 |  | The System must retrieve all defective families identified by Sherlock |
| F1.2 |  | The System shall abandon ASINs which are non-buyable and/or not part of variation family anymore |
| F1.3 |  | The System must retrieve existing catalog size attribute values for all Child ASINs for defective variation families |
| F1.4 |  | The System must retrieve ASINS with defect type as UOM duplicates, and normalize UOM by fetching values from normalization table (Appendix E) |
| F1.5 |  | The System must retrieve defect type as unit quantity duplicates, then normalize unit quantity values in size attribute by removing space and special characters |
| F1.6 |  | The System must retrieve ASINs with defect type as conversion duplicates, and source size attribute value from product image/description. If retrieved size value has the “UOM” of the family then use conversion table and normalize defective ASINs (Appendix F) |
| F1.7 |  | The System must check for ASIN duplicates with same UPC/EAN and merge them |
| F1.8 |  | The System must check for ASIN duplicates with different UPC/EAN and sideline them for manual validation |
| F1.9 |  | The System must update consistent and normalized size attribute values in production catalog |
| F2.0 |  | The System must provide information required for a human to perform Manual Correction when it fails to Auto Correct or where we want to perform an audit of an ASIN so it can learn from the corrections done |
| **F2** |  | **Inconsistent unit of measurement in Size** |
| F2.1 |  | The System must retrieve all defective families identified by Sherlock |
| F2.2 |  | The System shall abandon ASINs which non-buyable and/or not part of variation family anymore |
| F2.3 |  | The System must retrieve existing catalog size attribute values for all Child ASINs for defective variation families |
| F2.4 |  | The System must source unit of measurement values from image/product description for all defective child ASINs |
| F2.5 |  | The System must check of consistency of unit of measurement for all child ASINs in the family and if consistent values are found, update size attribute of defective ASINs with sourced values |
| F2.6 |  | If unit of measurement is not obtained or inconsistent, the system must sideline ASINs for manual validation/fixes. The System must provide information required for a human to perform Manual Correction when it fails to Auto Correct or where we want to perform an audit of an ASIN |
|  |  |  |

**Success criteria for ACE**

Should be able to fix all in scope defects with an accuracy level of >95%. Should be able to sideline 100% of defects that cannot be fixed for manual review.

**Appendix A -**   **Financial Impact**

**Consumables:** - We collected 1,500 BAND-A & B Consumables ASINs for a sample audit in which 1,250 ASINs had variations & 684 had variation errors. Assuming GVs, price, and other impacting factors are constant, a study on GLs with greater than 25% error rate (variation errors/ASINs with variations) in US MP (Grocery, Drugstore, and Beauty, GV Band A/B) revealed that non-error variation ASINs had average 7% higher shipped units per glance view than error variation ASINs for data quality errors.



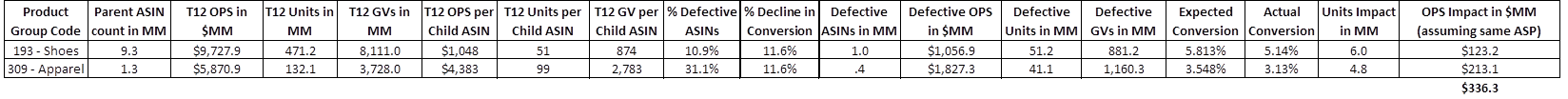
Leveraging above studies & assuming that similar lift is expected in shipped units per glance view for each product group in other market places, annualized impact calculation for GLs with error rate greater than 25% is:



For example, capturing 10% of this entitlement would lead to +124M in additional GMS.

Assumptions: ASP stays constant, GVs stay constant

**Softlines:** In order to measure the impact of ASIN defects within Softlines, the US SL ROI team manually audited a sample of 900 Parent ASINs that had at least 1 unit of sale recorded in 2018. These set of ASINs were then split into defective (ASINs that had any kind of relational or attribute defects) and non-defective (ASINs that had no defect recorded) buckets to measure the delta in unit conversion between the defective and non-defective bucket. A two sample t-test that was run on each of the 3 metrics (i) conversion per ASIN, (ii) glance views per ASIN and (iii) units per ASIN revealed a significant difference between the defective and non-defective ASINs at a 90% CI. Applying this % decline in conversion to the entire catalog (Shoes and Apparel) in the trailing 12 months showed a $336M impact to OPS. See table below for details.



**Appendix B – Conversion Impact**

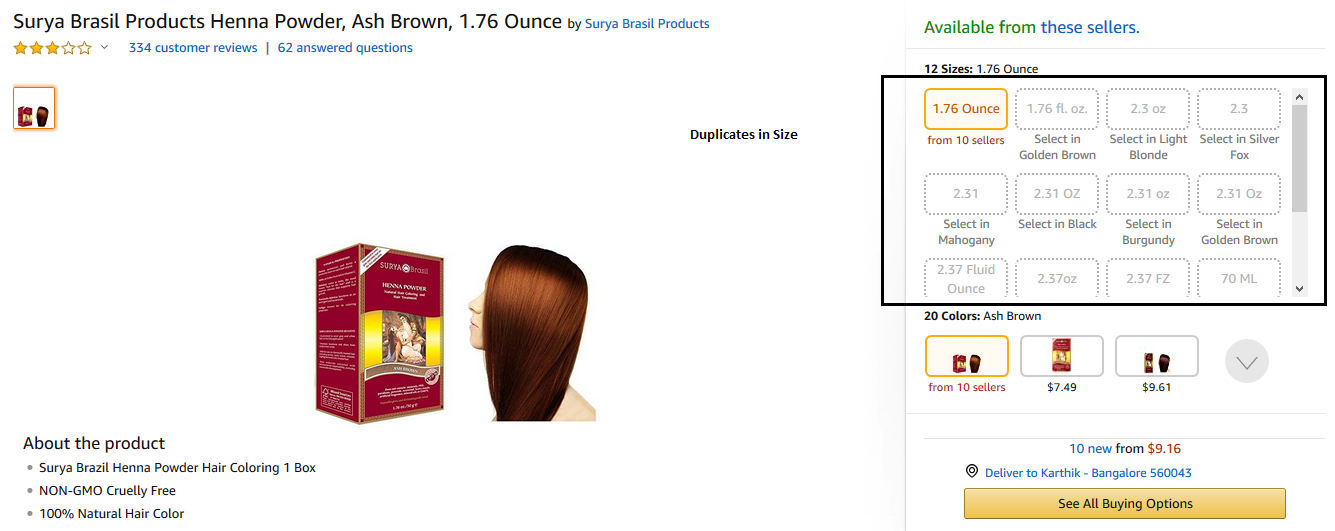
*Experiment on variation fixes for data quality errors resulted in +770 Bps improvement for test group when compared with similar homogeneous control group*



**Appendix C – Variation defects – Use case examples**

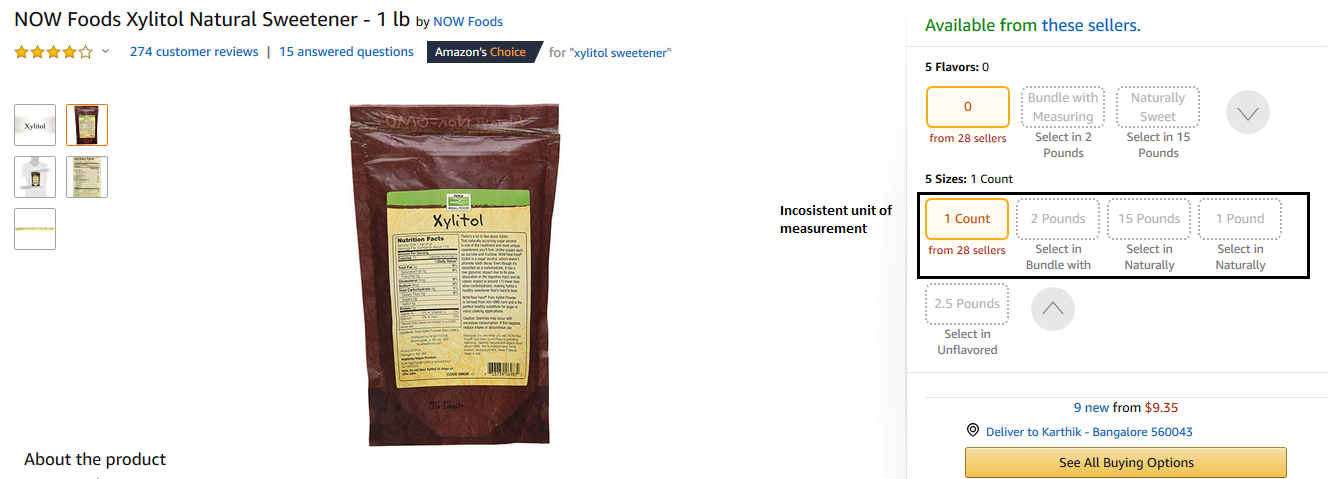
1. **Variation defect type: Duplicates in Size (**Size variation error where multiple of the same size are present in the same twister)

ASIN: [B0016AZA2G](https://www.amazon.com/dp/B001B6N2E8?th=1)



1. **Variation defect type: Inconsistent unit of measurement** (Size variation error where values have inconsistent units within the same twister)

ASIN: [B007729RBC](https://www.amazon.com/dp/B007729RBC?psc=1)

****

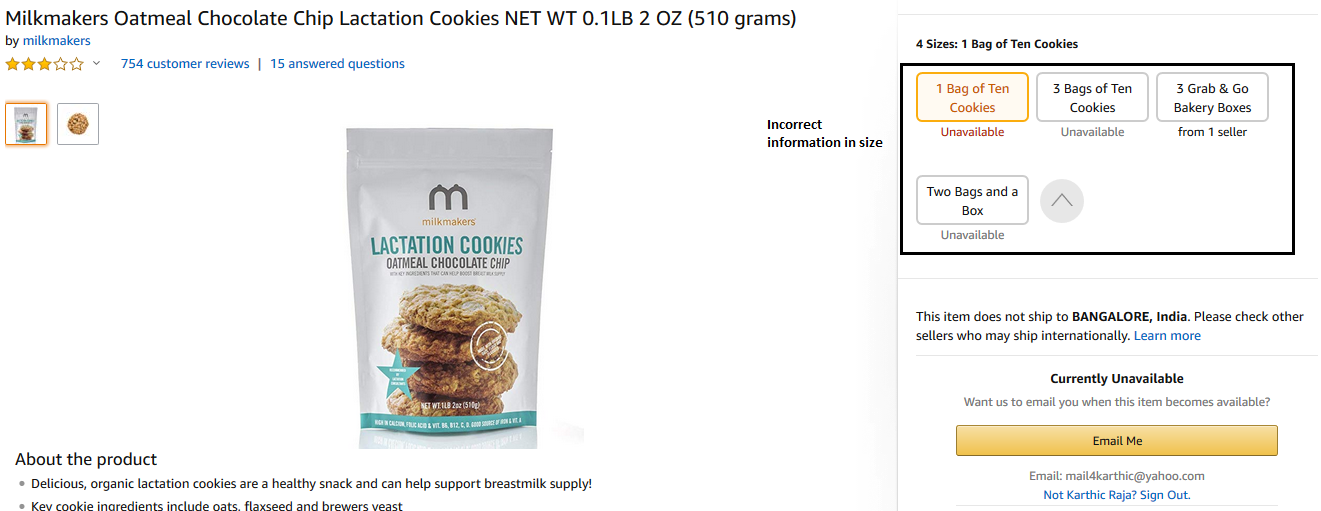
1. **Variation defect type: No unit of measurement in size** (Size variation error where value does not have a unit of measurement)

ASIN: [B002CQU4ZG](https://www.amazon.com/dp/B002CQU4ZG?psc=1)

****

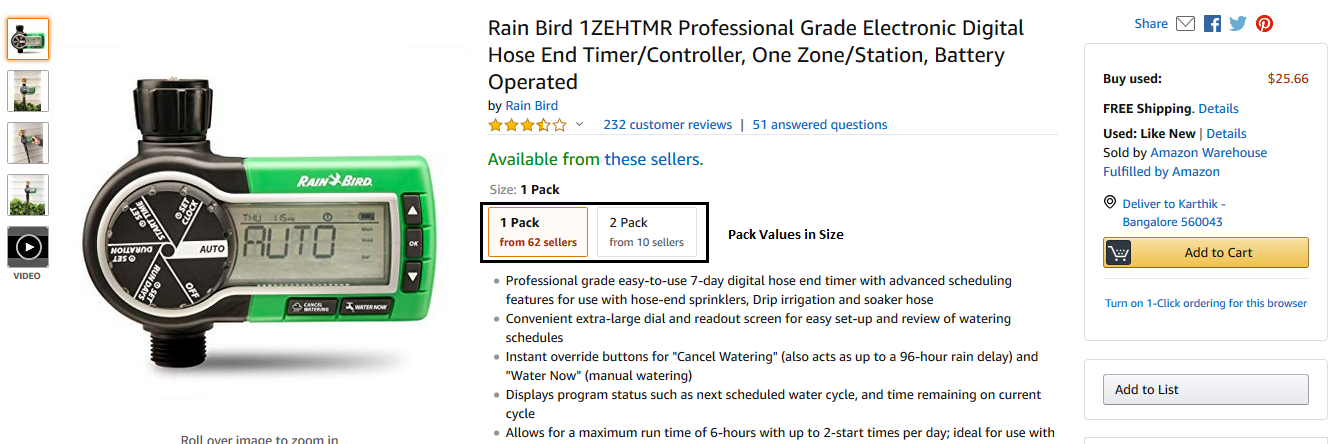
1. **Variation defect type: Incorrect information is Size** (Size variation error where non-size variation information is included in the size category)

ASIN: [B00A3J1RB4](https://www.amazon.com/dp/B00A3J1RB4?psc=1)



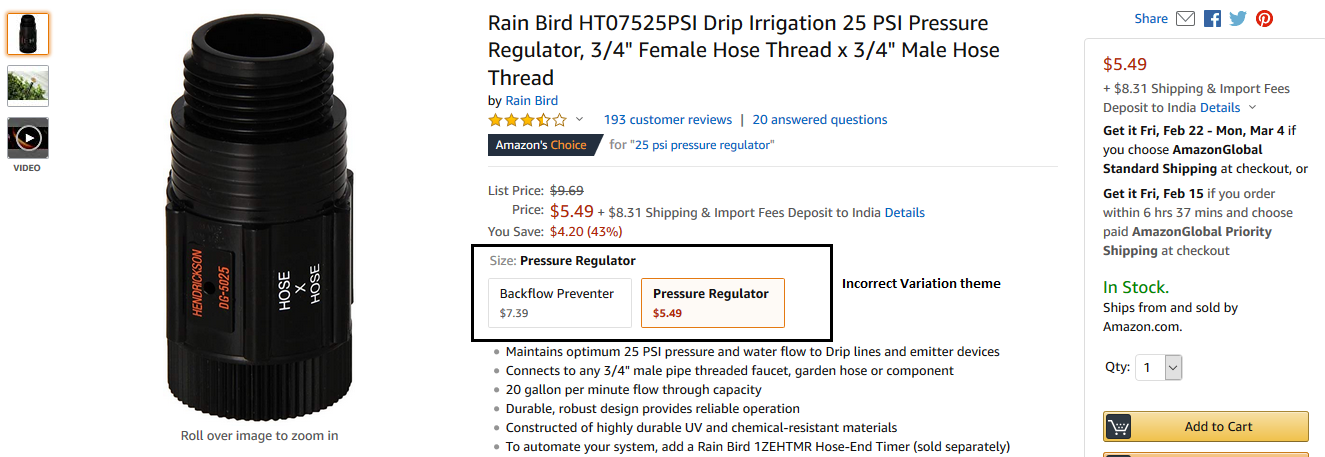
1. **Variation defect type: Pack values is Size** (Size variation error where value is a pack size and not a size)

ASIN:  [B006JZ514U](https://www.amazon.com/dp/B006JZ514U?psc=1)



1. **Variation defect type: Incorrect Variation theme** (Theme of parent/variation family is incorrect)

ASIN:  [B0049C5FZA](https://www.amazon.com/dp/B0049C5FZA?psc=1)



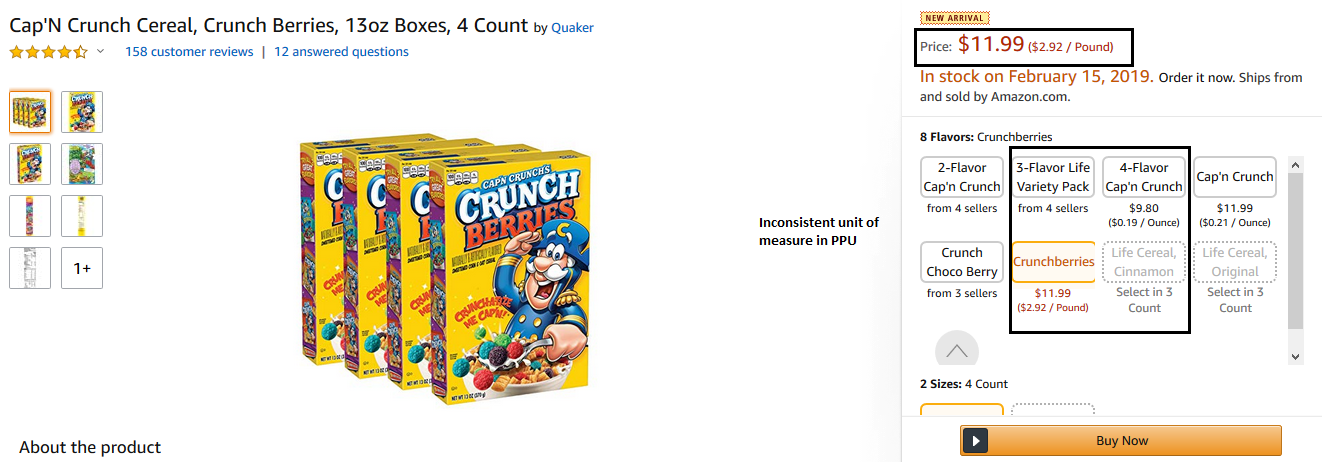
1. **Variation defect type: Incorrect variation family** ( Products groups within variation family are incorrect)

ASIN: [B07JVGTVZB](https://www.amazon.com/dp/B07JVGTVZB?psc=1)



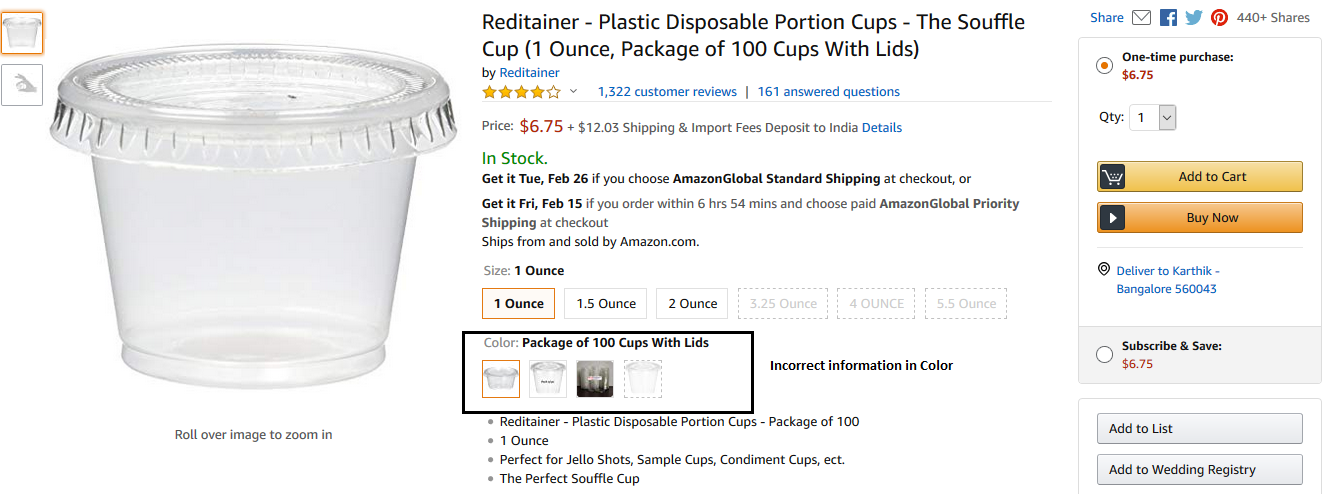
1. **Variation defect type: Missing and Inconsistent unit of measure in PPU** ( Unit of measure in PPU is either missing or inconsistent within the variation family)

ASIN: [B079Y5MNBW](https://www.amazon.com/dp/B079Y5MNBW?th=1)



1. **Variation defect type: Incorrect information in color** (Color attribute populated with style, size, scent, flavor and other irrelevant values)

ASIN: [B009VSFZ8Y](https://www.amazon.com/dp/B009VSFZ8Y?psc=1)



1. **Variation defect type: Incorrect information in other variating attributes** ( Irrelevant values in flavor, scent, style)

[ASIN: B01FO6WA9I](https://www.amazon.com/dp/B01FO6WA9I?psc=1)



**Appendix D – Variation program dashboard**

Link to actual file: <https://amazon.awsapps.com/workdocs/index.html#/folder/f3dbd3d9cb3bf5e5774b6e34a2e1cd0f07dd5d167b5348c32d6d8969ce4b4d54>





**Appendix E – UOM Normalization table**

Link to the actual file:

<https://amazon.awsapps.com/workdocs/index.html#/folder/eb63199d896e092d97ba7bdcc27b109c94b3187b05570f58d1c685d0844d76fd>

(A sample of normalization table)

|  |  |
| --- | --- |
| **Actual Value** | **Normalized Value** |
| LBS | Pound |
| lbs | Pound |
| lbs. | Pound |
| LB | Pound |
| lb | Pound |
| lb. | Pound |
| LB. | Pound |
| Pound | Pound |
| Pounds | Pound |
| pound | Pound |
| pounds | Pound |
| POUND | Pound |
| POUNDS | Pound |
| Ounce | Ounce |
| Ounces | Ounce |
| ounce | Ounce |
| ounces | Ounce |
| OUNCE | Ounce |
| OUNCES | Ounce |
| OZ | Ounce |
| oz | Ounce |
| oz. | Ounce |
| Oz | Ounce |
| Oz. | Ounce |
| oZ | Ounce |
| oZ. | Ounce |

**Appendix F – Conversion table**

Link to the actual file:

(A sample of conversion table)

|  |  |  |
| --- | --- | --- |
| UOM | Converted Value | Conversion |
| Pound | 16 Ounce | Pound to Ounce |
| Kilogram | 35.274 Ounce | Kilogram to Ounce |
| Gram | 0.03527 Ounce | Gram to Ounce |

**Appendix G – Size mapping table**

Link to the actual file: <https://amazon.awsapps.com/workdocs/index.html#/folder/15bd79527f5d8297b0f538282c8bdda788aa6cba00fd713060022873f18ae0eb>

(A sample of size mapping table)



**Appendix H – Pack Value strings**

|  |  |
| --- | --- |
| **Pack in Size`** | Pack, pack,(pack),(Pack), pck, Pck,( pack ),( ack),PACK,(pack,pack),packets,Packets,packet(s),Packet(s),pack of,Packs,packs,(Pack,(pack of,(Pack of |
| **Set in Size** | Set,set,(set),(Set),( set ),( set),SET,(set,set),set of,Sets,sets,(Set,(set of,(Set of,of |
| **Count in Size** | Count,count,(count),(Count),cnt,Cnt,( count ),( count),COUNT,(count,count), count of,Counts,counts,(Count,(count of,(Count of,of |

**Appendix I – Variation Use case prioritization**

\*\*Frequency (% defects) are based on random sample audit of 1500 ASINs across US Consumables

\*\*Business impact has been determined based on conversion impact measurements through coffee pilot

\*\*Action-ability based on current status of the program

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No** | **Variation Defect type** | **Frequency (% defects)** | **Business impact** | **Action ability** | **Priority** |
| 1 | Duplicates in size | 3% | High | Yes | 1 |
| 2 | Inconsistent unit of measure in size | 9% | High | Yes | 1 |
| 3 | No unit of measurement in size | 1% | High | Yes | 1 |
| 4 | Incorrect information is size | 11% | High | Yes | 1 |
| 5 | Pack values in size | 25% | Low | No | 4 |
| 6 | Pack values in other variating attributes | 2% | Low | No | 4 |
| 7 | Incorrect Variation theme | 6% | Low | Yes | 4 |
| 8 | Missing PPU | 18% | High | Yes | 2 |
| 9 | Incorrect unit of measure in PPU | High | Yes | 2 |
| 10 | Inconsistent unit of measure in PPU | High | Yes | 2 |
| 11 | Incorrect information in color | 7% | Med | Yes | 3 |
| 12 | Incorrect information in other variating attributes | 3% | Med | Yes | 5 |
| 13 | Incorrect variation family | 2% | High | Yes | 1 |
| 14 | Other out of scope of defects | 13% | NA | NA | NA |

**Appendix J – SOP for Variation Corrections**

Link to the actual file:

<https://amazon.awsapps.com/workdocs/index.html#/folder/eb63199d896e092d97ba7bdcc27b109c94b3187b05570f58d1c685d0844d76fd>

**Variation fixes SOP for Duplicates in Size**

* Retrieve all ASINs within a family identified as duplicates by Sherlock
* Retrieve existing catalog size attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* Check If the defect type marked by Sherlock is UOM duplicate
  + De-string only UOM from size attribute
  + Normalize UOM by matching with Normalization table
  + If value not found in the normalization table, retrieve UOM from product image/description and then normalize UOM
* Check If the defect type marked by Sherlock is Unit Count duplicate
  + Normalize unit count by removing spaces & special characters
  + Round to one decimal value
* Check If the defect type marked by Sherlock is conversion duplicate
  + Retrieve size attribute values from product image/description
  + Check all ASINs within the family and arrive at the "UOM" for the family
  + Check retrieved size value has "UOM" of the family
    - Use conversion table and normalize non-standard ASINs
    - Round to one decimal value
* Check if the normalized size value is greater than 24 characters
  + If greater than 24 characters, sideline for manual validation
* Check for consistency of UOM, across the family before correction
* Check Size value contribution,
  + If from retail
    - Update validated size attribute value in production catalog through selection central
  + If from 3P, and if precedence greater than 49.9
    - Leverage data augmenter account to update validated size attribute values in production catalog
  + If from 3P, and if precedence less than 49.9
    - Update validated size attribute value in production catalog through SPOT
* If UPC/EAN of duplicate ASINs are same
  + Use Darwin to merge ASINs
* If UPC/EAN of duplicate ASINs are different
  + Check catalog attributes - title, unit count, unit count type, variation attributes, product description, image to validate both are same products
  + Perform bin check only when more clarity is needed
  + If asins are confirmed as same products
    - If retail/FBA ASINs
      * Check on-hand inventory/on-order inventory
      * If both ASINS have inventory
        + Remove ASIN with lower OPS out of the variation family
      * If only one ASIN has inventory
        + Suppress the ASIN without inventory using data augmenter
    - If 3P ASINs
      * Merge ASINS through Darwin tool
  + If asins are confirmed as different products
    - Validate product image and description of all child asins to check the variating factor
    - If retail Parent
      * Use selection central to change the new variation theme with size and other variating factor
      * Check attribute contribution of each child ASINs
      * If from retail
        + Update the relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Update the relevant variating attribute value in production catalog through data augmenter account
    - If 3P Parent
      * Use Vermont/Darwin tool to create a new parent with new variation theme (Size and other variating attribute)
      * Align all Child ASINs to the new parent ASIN
      * Delete the existing parent ASIN
      * Check attribute contribution of each child ASINs
        + If from retail

Update the relevant variating attribute value in production catalog through selection central

* + - * + If from 3P, and if precedence greater than 49.9

Update the relevant variating attribute value in production catalog through SPOT

* + - * + If from 3P, and if precedence less than 49.9

Update the relevant variating attribute value in production catalog through data augmenter account

**Variation fixes SOP for Incorrect UOM in Size**

* Retrieve all ASINs within a family identified as Incorrect UOM in size by Sherlock
* Retrieve existing catalog Size attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* Retrieve size attribute values from product image/description
  + Check all ASINs within the family and arrive at the "UOM" for the family by validating with the size mapping table
  + If retrieved values doesn’t have a UOM, work with vendor/seller to get the right UOM values
* Normalize all UOMs by matching with normalization table
* Check for consistency of UOM, across the family before correction
  + If the UOM of the family not consistent, sideline for manual validation
* Check if the normalized size value is greater than 24 characters
  + If greater than 24 characters, sideline for manual validation
* Check Size value contribution,
  + If from retail
    - Update validated size attribute value in production catalog through selection central
  + If from 3P, and if precedence greater than 49.9
    - Leverage data augmenter account to update validated size attribute values in production catalog
  + If from 3P, and if precedence less than 49.9
    - Update validated size attribute value in production catalog through SPOT

**Variation fixes SOP for No unit of measurement in size**

* Retrieve all ASINs within a family identified as No unit of measurement in size by Sherlock
* Retrieve existing catalog Size attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* Retrieve size attribute values from product image/description
  + Normalize retrieved values by validating with normalization table
* Check for consistency of UOM, across the family before auto correction
  + If non consistent sideline for manual validation
* Check if the normalized size value is greater than 24 characters
  + If greater than 24 characters, sideline for manual validation
* Check Size value contribution,
  + If from retail
    - Update validated size attribute value in production catalog through selection central
  + If from 3P, and if precedence greater than 49.9
    - Leverage data augmenter account to update validated size attribute values in production catalog
  + If from 3P, and if precedence less than 49.9
    - Update validated size attribute value in production catalog through SPOT

**Variation fixes SOP for Incorrect information in Size**

* Retrieve all ASINs within a family identified as Incorrect information in size by Sherlock
* Retrieve existing catalog size attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* If size attribute has only the size value
  + Retrieve size attribute values from product image/description
  + Normalize retrieved values by validating with normalization table
  + Check for consistency of UOM, across the family before auto correction
    - If non consistent sideline for manual validation
  + Check if the normalized size value is greater than 24 characters
    - If greater than 24 characters, sideline for manual validation
  + Check Size value contribution,
    - If from retail
      * Update validated size attribute value in production catalog through selection central
    - If from 3P, and if precedence greater than 49.9
      * Leverage data augmenter account to update validated size attribute values in production catalog
    - If from 3P, and if precedence less than 49.9
      * Update validated size attribute value in production catalog through SPOT
* If size has only other variating attribute values (Color/Flavor/Scent/Style)
  + Validate product image and description of all child asins to check the variating factor
  + If retail Parent
    - Use selection central to change the new variation theme
    - Check attribute contribution of each child ASINs
      * If from retail
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through data augmenter account
  + If 3P Parent
    - Use Vermont/Darwin tool to create a new parent with new variation theme
    - Align all Child ASINs to the new parent ASIN
    - Delete the existing parent ASIN
    - Check attribute contribution of each child ASINs
      * If from retail
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Remove attribute values from size and update the relevant variating attribute value in production catalog through data augmenter account
* If size has both size and other variating attribute values (Color/Flavor/Scent/Style)
  + Validate product image and description of all child asins to check the variating factors
  + If retail Parent
    - Use selection central to change the new variation theme with two dimensions (size and other variating attribute)
    - Split both Size and other variating attribute values
    - For all child ASINs, Retrieve size attribute values from product image/description
    - Normalize retrieved values by validating with normalization table
    - Check for consistency of UOM, across the family before auto correction
      * If non consistent sideline for manual validation
    - Check if the normalized size value is greater than 24 characters
      * If greater than 24 characters, sideline for manual validation
    - Check Size value contribution,
      * If from retail
        + Update validated size attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Leverage data augmenter account to update validated size attribute values in production catalog
      * If from 3P, and if precedence less than 49.9
        + Update validated size attribute value in production catalog through SPOT
    - For all child ASINs, populate other variating attribute details in the new variation attribute created
    - Check attribute contribution of each child ASINs
      * If from retail
        + Update relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Update the relevant variating attribute value in production catalog through data augmenter account
  + If 3P Parent
    - Use Vermont/Darwin tool to create a new parent with two dimensional variation theme (Size and other variating attribute)
    - Align all Child ASINs to the new parent ASIN
    - Delete the existing parent ASIN
    - Split both Size and other variating attribute values
    - For all child ASINs, Retrieve size attribute values from product image/description
    - Normalize retrieved values by validating with normalization table
    - Check for consistency of UOM, across the family before auto correction
      * If non consistent sideline for manual validation
    - Check if the normalized size value is greater than 24 characters
      * If greater than 24 characters, sideline for manual validation
    - Check Size value contribution,
      * If from retail
        + Update validated size attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Leverage data augmenter account to update validated size attribute values in production catalog
      * If from 3P, and if precedence less than 49.9
        + Update validated size attribute value in production catalog through SPOT
    - For all child ASINs, populate other variating attribute details in the new variation attribute created
    - Check attribute contribution of each child ASINs
      * If from retail
        + Update relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Update the relevant variating attribute value in production catalog through data augmenter account
* If size has pack/quantity values
  + Solution to be defined

**Variation fixes SOP for Pack Values in Size and other variation attribute values**

* + Solution to be defined

**Variation fixes SOP for Missing PPU**

* Retrieve all ASINs within a family identified as Missing PPU
* Retrieve existing catalog unit count, unit count type attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* Retrieve unit count, unit count type attribute values from product image/description
  + Normalize retrieved values by validating with normalization table
* Check for consistency of unit count type, across the family before correction
  + If non consistent sideline for manual validation
* Check unit count & unit count type value contribution,
  + If from retail
    - Update validated unit count and unit count type attribute value in production catalog through selection central
  + If from 3P, and if precedence greater than 49.9
    - Leverage data augmenter account to update validated unit count and unit count type attribute values in production catalog
  + If from 3P, and if precedence less than 49.9
    - Update validated unit count and unit count type attribute values in production catalog through SPOT

**Variation fixes SOP for Incorrect/Inconsistent unit of measure in PPU**

* Retrieve all ASINs within a family identified as Incorrect/Inconsistent unit of measure in PPU
* Retrieve existing catalog unit count type attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* Retrieve unit of measure values from product image/description
  + Check all ASINs within the family and arrive at the "UOM" for the family by validating with the size mapping table
  + If retrieved values doesn’t have a UOM, work with vendor/seller to get the right UOM values
* Normalize all UOMs by matching with normalization table
* Check for consistency of UOM, across the family before correction
  + If inconsistent sideline for manual validation
* Check unit count type value contribution,
  + If from retail
    - Update validated unit count type attribute value in production catalog through selection central
  + If from 3P, and if precedence greater than 49.9
    - Leverage data augmenter account to update validated unit count type attribute value in production catalog
  + If from 3P, and if precedence less than 49.9
    - Update validated unit count type attribute values in production catalog through SPOT

**Variation fixes SOP for Incorrect information in Color**

* Retrieve all ASINs within a family identified as Incorrect information in color by Sherlock
* Retrieve existing catalog color attribute values for the list of input ASINs
* Check ASIN is buyable
* Check ASIN part of variation family
* If color attribute has only the color value
  + Check retrieved color value with the color mapping table
  + Normalize the attribute value using the color mapping table
  + Check if the normalized color value is greater than 24 characters
    - If greater than 24 characters, sideline for manual validation
  + Check color attribute value contribution,
    - If from retail
      * Update validated color attribute value in production catalog through selection central
    - If from 3P, and if precedence greater than 49.9
      * Leverage data augmenter account to update validated color attribute values in production catalog
    - If from 3P, and if precedence less than 49.9
      * Update validated color attribute value in production catalog through SPOT
* If Color has only other variating attribute values (Size/Flavor/Scent/Style)
  + Validate product image and description of all child asins to check the variating factor
  + If retail Parent
    - Use selection central to change the new variation theme
    - Check attribute contribution of each child ASINs
      * If from retail
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through data augmenter account
  + If 3P Parent
    - Use Vermont/Darwin tool to create a new parent with new variation theme
    - Align all Child ASINs to the new parent ASIN
    - Delete the existing parent ASIN
    - Check attribute contribution of each child ASINs
      * If from retail
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Remove attribute values from color and update the relevant variating attribute value in production catalog through data augmenter account
* If color has both size and other variating attribute values (Size/Flavor/Scent/Style)
  + Validate product image and description of all child asins to check the variating factors
  + If retail Parent
    - Use selection central to change the new variation theme with two dimensions (color and other variating attribute)
    - Split both color and other variating attribute values
    - Check retrieved color value with the color mapping table
    - Normalize the attribute value using the color mapping table
    - Check if the normalized color value is greater than 24 characters
      * If greater than 24 characters, sideline for manual validation
    - Check color value contribution,
      * If from retail
        + Update validated color attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Leverage data augmenter account to update validated color attribute values in production catalog
      * If from 3P, and if precedence less than 49.9
        + Update validated color attribute value in production catalog through SPOT
    - For all child ASINs, populate other variating attribute details in the new variation attribute created
    - Check attribute contribution of each child ASINs
      * If from retail
        + Update relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Update the relevant variating attribute value in production catalog through data augmenter account
  + If 3P Parent
    - Use Vermont/Darwin tool to create a new parent with two dimensional variation theme (Color and other variating attribute)
    - Align all Child ASINs to the new parent ASIN
    - Delete the existing parent ASIN
    - Split both Color and other variating attribute values
    - Check retrieved color value with the color mapping table
    - Normalize the attribute value using the color mapping table
    - Check if the normalized color value is greater than 24 characters
      * If greater than 24 characters, sideline for manual validation
    - Check Color value contribution,
      * If from retail
        + Update validated color attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Leverage data augmenter account to update validated color attribute values in production catalog
      * If from 3P, and if precedence less than 49.9
        + Update validated color attribute value in production catalog through SPOT
    - For all child ASINs, populate other variating attribute details in the new variation attribute created
    - Check attribute contribution of each child ASINs
      * If from retail
        + Update relevant variating attribute value in production catalog through selection central
      * If from 3P, and if precedence greater than 49.9
        + Update the relevant variating attribute value in production catalog through SPOT
      * If from 3P, and if precedence less than 49.9
        + Update the relevant variating attribute value in production catalog through data augmenter account
* If color has pack/quantity values
  + Solution to be defined yet