International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V72I5P123

Volume 72 Issue 5, 188-197, May 2024

Original Article

Predictive Analysis of Smart Spray Solutions

Jeshwanth Kundem[1], Michelle Magdziarz[2], Ger Vue[3], Chia Yathortou[4]

[1,2,3,4]Opus College of Business, University of St. Thomas, Minnesota, USA.
(affiliation when work was conducted)

[1]Corresponding Author : jeshwanth.kundem@outlook.com

Received: 25 March 2024 Revised: 30 April 2024 Accepted: 11 May 2024 Published: 22 May 2024

Abstract *- The digitization of products is considered the fourth industrial revolution. The spray paint applicator manufacturers are presently assessing the market to digitize their products through multiple smart solutions. A confined market survey on various aspects of digitization has been performed, on which descriptive and statistical analysis has been conducted. Results indicate that most of the features are not correlated with most of the demographic/ user profile variables. However statistical analysis suggests that the experience level of participants has significance in dictating the interest level in features. Most of the participants are interested in features that come under the notification category.*

Keywords *- Paint sprayers, Predictive analysis, Descriptive analysis, Digitization, Airless.*

1. Introduction

Smart products are vital in the present market for the survival of the business. With computer technology asymptotically approaching a reliable stage in terms of digitization, engineering companies are advancing their products through the implementation of various smart devices. Smart devices are connected products/ things that are embedded with processors, sensors, connectivity, etc., that allow data to be exchanged between the product and its environment, manufacturer, operators, and other products.

Smart products not only play a role in terms of productivity but also in the engineering simulation world, where engineers are mimicking the product on the field through real-time data. This real-time data collection and creation of a digital twin, has boosted the engineer's creativity in troubleshooting, maintenance, and creation of the next generation of products through understanding the customer effectively. Spray paint applicators are no different.

The digitization of spray paint applicators comes in various forms, including but not limited to real-time data on applicator performance, maintenance tracking, spray characteristics, remote control of the sprayer, etc. With multiple constructs to study, a survey has been formulated that describes several aspects of types of digitization for spray paint applicators. Based on the survey results various descriptive and quantitative analyses have been performed. A few tools used for the analysis were Tableau and Minitab.

As this is an industrial-based analysis, there are no articles that describe the research. However, various similar articles have been put in the literature, for the understanding of the reader.

2. Survey

To investigate the present market for digital paint applicators, a survey has been conducted. The census of interest is anyone who is/ was a professional painter. As it is a survey, the sampling method would be a non-probability sampling method – a voluntary sample. Since the email database is all the professional painters, it is expected that the survey takers have an interest in the topics of the survey. The survey takers establish the sample rather than the survey organizer.

A survey questionnaire has been developed and distributed to the recipients in the email database through Qualtrics. It included the gathering of measurement variables related to the concepts of the user profile, smart features, and demographics. Throughout the survey, the questions were formulated in the terminology that a professional paint sprayer would understand.

For the reader's reference, a professional paint sprayer is a device that contains a pump with mechanical parts, a hose, and a gun with a spray nozzle (Figure 1). The following subsections shall describe the questions in the survey and the corresponding variable name that shall be employed in further sections of the article.

2.1 User Profile

Under this construct, questions related to the sprayer's profession, experience, type of product usage, and usage rate have been formulated.

Fig. 1 Typical paint sprayer (from the internet)

- Have you used a professional paint sprayer at least once within the past 90 days?
  - Knowing whether the survey taker is/ was a user of a paint sprayer is essential for accurate estimation of the market.
  - Variable: PaintedYesNo (P)

- I classify myself as a:
  - Survey taker's profession
  - Variable: User Type (UT)

- How many years have you worked in the painting industry?
  - Survey taker's experience
  - Variable: Experience (E)

- How many gallons of paint (all coatings) do you spray in a typical week?
  - Survey taker's usage rate
  - Variable: Usage Rate (UR)

- What method(s) of application do you use to apply coatings?
  - Survey taker's choice of various products of professional painting equipment.
  - Variables: Brush Roll (BR), Electric Airless Sprayer (EAS), Air Assisted Sprayer (AAS), Gas Airless Sprayer (GAS), Line Striper (LS), HVLP Sprayer (HVLP)

2.2. Smart Features

Smart Feature variables are the core of the survey and the analysis. The importance of various smart features, including but not limited to connectivity, real-time quantification, remote access, etc., have been referenced in this section.

- A feature that would enable the pump to know the true tip size and notify you when the tip needs to be replaced.
  - Tip (nozzle) gets worn off during extended use of the paint sprayer. Knowing the true tip size is essential for the sprayer to have better performance on the job.
  - Variable: Tip Replacement Notification (TRN)

- A feature that would enable you to have the data from the pump, tip, and coating collected, stored, and available remotely.
  - Data collection related to the characteristics of the pump
  - Variable: Remote Data Collection (RD)

- A feature that would enable you to track and store environmental data (temperature, humidity, wind speed, etc.) to help better understand job-site dynamics
  - Understanding the environmental conditions shall help the sprayer perform well on his job and also minimize overspray, which is a health concern.
  - Variable: Environmental Conditions (EC)

- A feature that would enable you to see the point of view spraying recommendations (proper distance to the wall, hand speed, overlap) via a device on the spray gun
  - Knowing the theoretical/ recommended way of spraying to optimize the performance of the coating better
  - Variable: Point of View (POV)

- A feature that would enable you to lock/ disable your pump(s) remotely.
  - Ability to secure their equipment
  - Variable: Remote Lock (RL)

- A feature that would enable you to scan a coating's barcode, and recommend set pressure and tip size.
  - Different coating (paint) materials have different fluid properties. The question addresses usage recommendations based on those properties.
  - Variable: UPC Code (UPC)

- A feature that would enable you to update software remotely to optimize your pump's performance
  - Similar to Windows updates or app updates on the phone, the pump has software embedded, that routinely is improved through bug fixes and the addition of new features.
  - Variable: Remote Software Update (RS)

- A feature that would enable you to monitor your pump's mechanical parts and provide service notifications
  - Mechanical parts wear off over time. Keeping an eye on their life and providing regular notifications.
  - Variable: Service Notifications (SN)

- A feature that would enable you to photograph your spray pattern and generate recommendations on set pressure, tip size, clean filters, etc., for better spray pattern results
  - A paint job will be judged by the final looks of the coating. This feature shall conduct an image

processing analysis on the initial spray of the job to better recommend the pump characteristics.
  - Variable: Spray Pattern Characteristics (SPC)

- A feature that would enable you to know where your pump(s) is at all times via GPS monitoring and tracking.
  - Information on the physical location of the pump, when the owner owns multiple pumps or lends his pump to other parties.
  - Variable: GPS Monitoring (GPS)

- A feature that would enable you to recognize and order pump parts and accessories, instantaneously by capturing a picture on a camera.
  - Through Augmented reality (AR), the user can locate the mechanical parts of the pump.
  - Variable: Augmented Reality of Parts (AR)

- A feature that would enable you to determine in real-time whether you have enough material to complete a job via a tool that continually calculates the amount of coating required.
  - Information on the amount of paint applied concerning the area coated to manage the available paint for the rest of the job.
  - Variable: Real-time paint calculations (RP)

- A feature that would enable you to have the pump automatically recognize the coating type, set the pressure, and recommend the appropriate tip.
  - Rather than scanning the barcode of the coating, this feature will recognize the coating type by itself.
  - Variable: Appropriate tip size recommendation (TSR)

- A feature that would enable you to create a digital invoice based on data retrieved from the pump (spraying hours, exact gallons sprayed, linear feet, etc.)
  - Having accurate information on the job conducted, one would be able to bill their customer accurately.
  - Variable: Digital Invoice (DI)

2.3. Demographics

For a market survey, having an understanding of the user's location and age group is important regarding focusing the products on the corresponding personnel.

- What is your age?
  - Age group, as the smart features are oriented with new technological advances.
  - Variable: Age (A)

- I presently live in (State/ Provinces):
  - Location of the user – US states and Canadian provinces
  - Variable: State (S)

3. Methodology

3.1. Data Cleaning and Descriptive Analysis

The survey has been online for 10 days. During this period, 283 users have taken the survey. Among them, 30 did not finish the survey. Due to that, the reduction of the dataset to 253 has been employed. Within the 253 samples, 11 survey takers answered NO for the question on the variable PaintedYesNo. As knowing the opinion of present sprayers is more vital for a new product introduction, the 11 data points have been removed from the dataset. The final dataset for the analysis contains N = 242.

As the features are of primary importance, multiple descriptive analyses have been performed initially, addressing feature importance concerning the type of pump, age group, user type, usage rate, and location.

3.2. Statistical Analysis

For this analysis, data reduction techniques have been employed for the variables in the Features construct. Due to a lack of independent and dependent variables, it is determined that either Primary Component Analysis or Exploratory Factor Analysis would be used to analyze the collected data. Mathematically, PCA is a simple form of factor analysis where rotations and transformations have not been accounted for. EFA is approached with prior knowledge on the belief that which variables may be similar and, therefore should load on a single factor. EFA considers interpretability between the variables, while PCA does not.

Based on the correlations between the factors (explained in the section Results and Discussions), it is decided that EFA would be ideal for data reduction in the present problem, and further analysis has been performed based on the factors derived.

4. Results

4.1. Descriptive Analysis

Initially, descriptive analysis has been conducted to have a visual understanding of the dataset. Things that were considered are how the features vary with the type of pump, age group, usage type, usage rate, and location.

4.1.1. Variation of Features with the Type of Pump

Here a graphical analysis has been performed to determine the top 3 features of interest by pump type.

Initially, for the pump type BR, the participants, with answers "never", "occasionally", and "rarely" for the type of application coating, have been removed from the dataset. Then "frequently" and "always" were combined into one category. After this, an average score for each feature is given to determine what is most important for that type of pump by recoding the importance level (Table 1). This process has been repeated independently for the rest of the pump types (EAS, AAS, GAS, HVLP, LS) and the results have been presented in Figure 2.

The most important features for most of the pump types are Service Notifications (SN) and Tip Replacement Notification (TRN). Both these features are in the top 3 for all the pump types

Table 1. Recoding of importance levels

Not at all important	1
Slightly important	2
Moderately important	3
Important	4
Extremely important	5

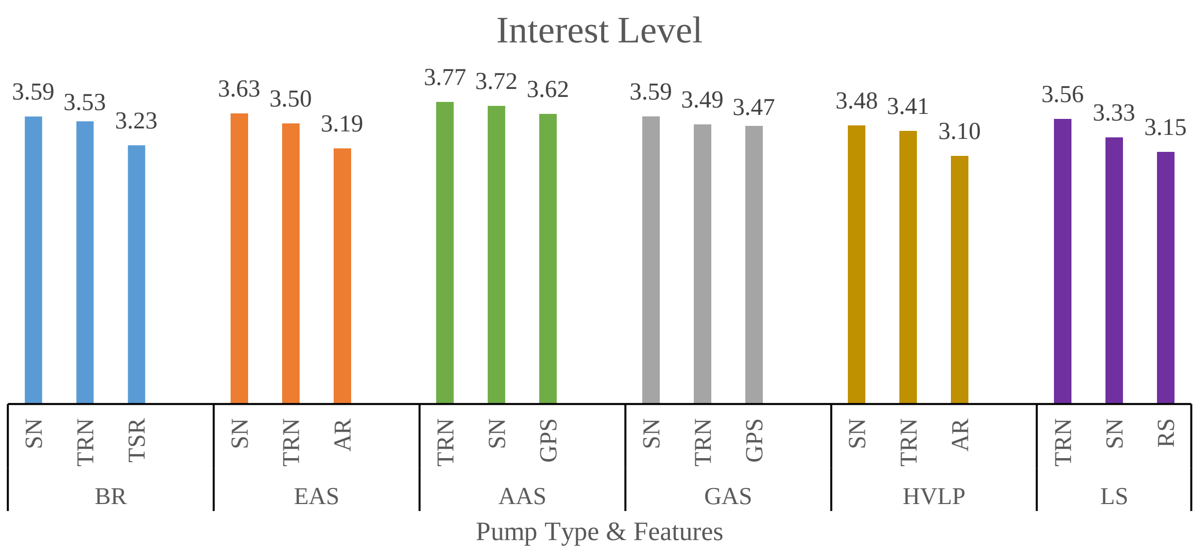

Fig. 2 Variation of features with a pump type

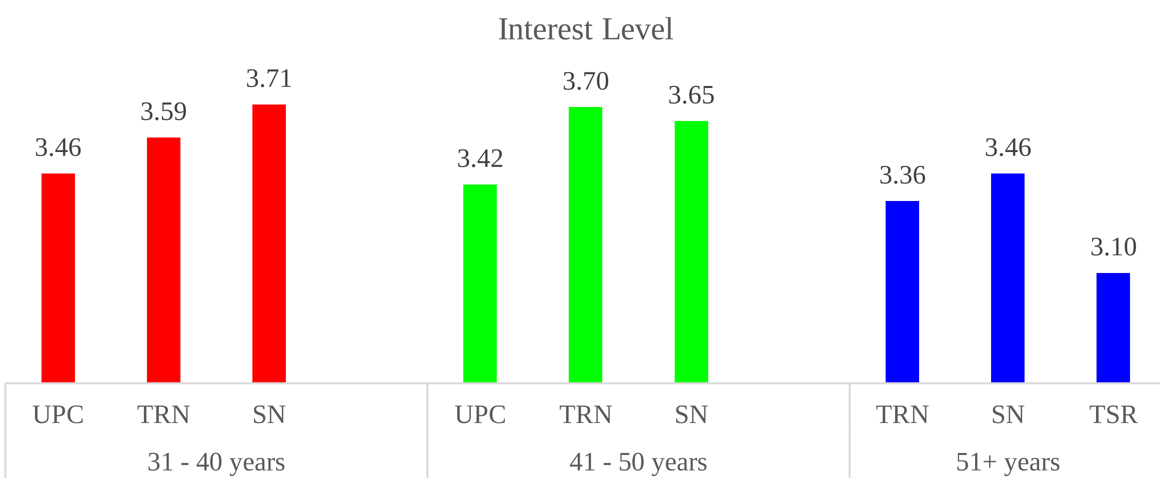

Fig. 3 Interest Level variation of top 3 features with age group

4.2.2. Variation of Features with Age Group

The average scores of the features have been taken into account, and the top 3 features for each group are presented here. In the dataset, it is observed that there are only 2 participants with age less than 30 years. So, the categories less than 18 years, 18 – 25 years, and 26 – 30 years have been removed from this descriptive analysis. Figure 3 indicates the top 3 features needed for the rest of the age groups. Based on the graph (Figure 3) for the age groups 31 – 40 years and 41 – 50 years, UPC, TRN, and SN are the top 3 important features. While for the 51+ years age group, UPC is not of importance. However, it is to be noted that, irrespective of the age group, Service Notification (SN) and Tip Size Recommendation (TSR) are much-needed features.

In Figure 4, a circular plot of all the features, colored with age group, is shown for the reader's visual interpretation.

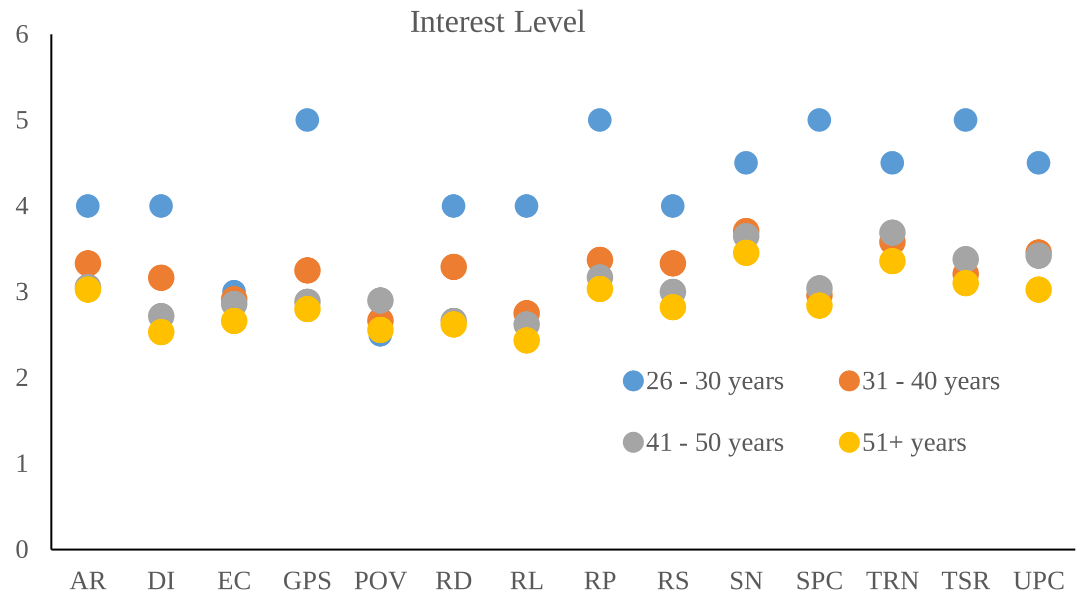

Fig. 4 Variation of all features with age group

4.2.3. Variation of Features with User Type

Similar to the analysis before, the average scores of the features are taken into account. The top 3 features needed by each user type have been presented in Figure 5, and the average scores of all the features with user type have been presented in Figure 6.

TRN and TSR seem to be the most important features for every user type. Surprisingly no one is interested in the Remote Lock (RL) feature even though some participants in the survey are owners of the painting company.

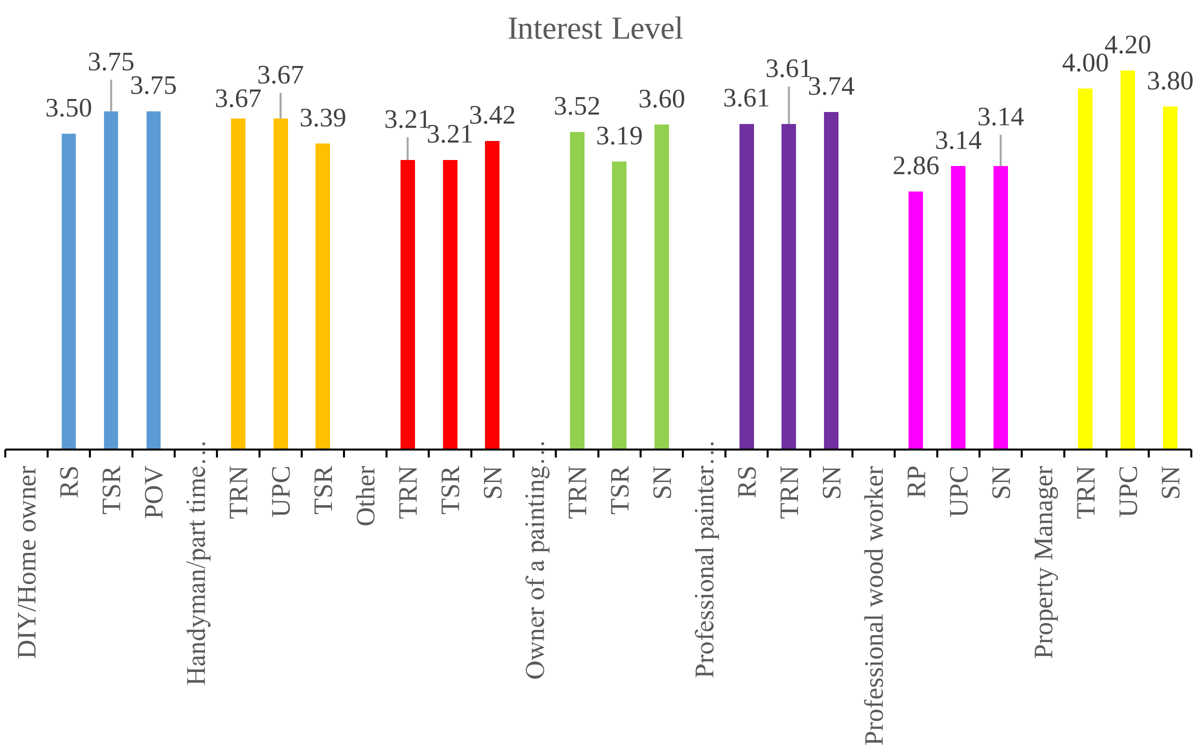

Fig. 5 Variation of top 3 features with user type

	DIY/Home owner	Handyman/Part time painter	Owner of a painting company	Professional painter (non-owner)	Professional wood worker	Property Manager	Other
AR	2.75	3.06	3.08	3.35	2.71	2.60	3.00
DI	2.25	2.50	2.67	2.96	1.57	3.20	2.68
EC	3.00	2.33	2.74	2.91	2.29	3.20	3.00
GPS	2.50	2.28	3.01	3.17	1.86	2.60	2.58
POV	3.75	2.89	2.63	2.52	2.57	3.40	2.63
RD	3.00	2.28	2.67	3.13	2.43	3.60	2.74
RL	2.75	2.39	2.59	2.74	1.43	2.60	2.26
RP	3.25	3.17	3.14	3.22	2.86	3.40	2.79
RS	3.50	2.83	2.85	3.61	2.29	3.60	2.89
SN	2.75	3.22	3.60	3.74	3.14	3.80	3.42
SPC	3.25	3.06	2.84	3.26	2.57	4.20	2.95
TRN	3.25	3.67	3.52	3.61	2.43	4.00	3.21
TSR	3.75	3.39	3.19	3.17	2.86	3.60	3.21
UPC	3.50	3.67	3.10	3.35	3.14	4.20	3.16

Fig. 6 Variation of all features with user type

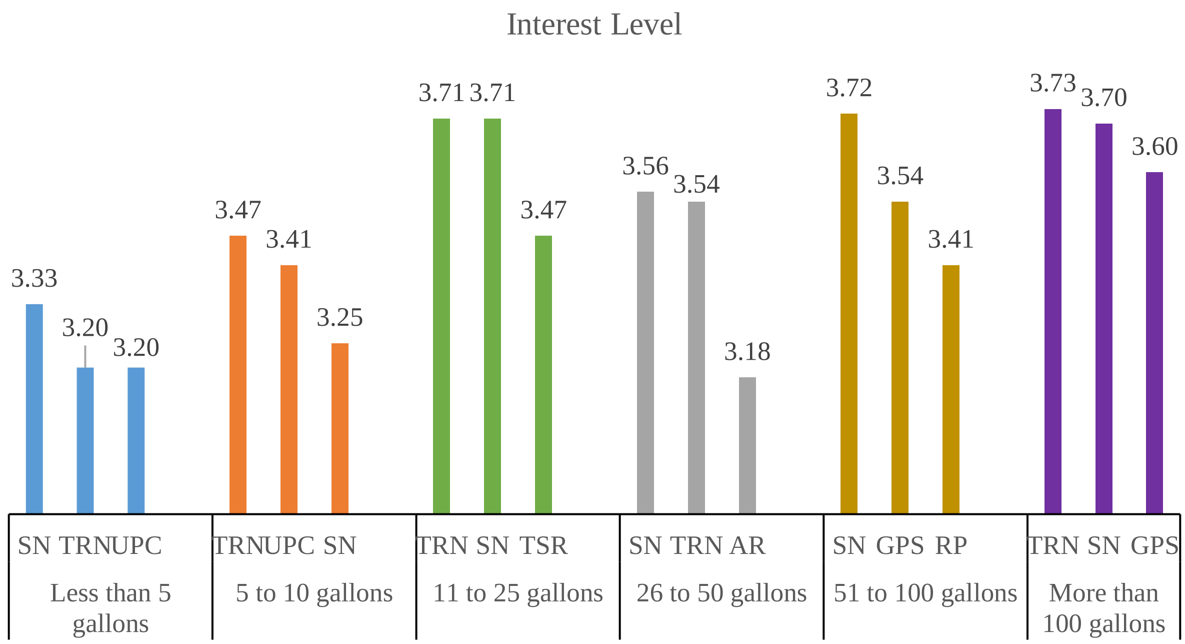

Fig. 7 Variation of features with user rate

4.2.4. Variation of Features with User Rate

Similar to the previous analysis, average scores are considered for the features to determine the top 3 features of interest based on usage rate (Figure 7). Based on the graph, the most important features are still Service Notifications (SN) and Tip Replace Notification (TRN).

4.2.5. Variation of Features with Location

Table 2 summarizes what feature(s) each state is most interested in and what pump is most used for coating by state. The table only includes states with 6 or more responses. The features were recoded as explained before. To determine the pump most used for coating by state, a similar recording has been conducted with 1 being "never" and 5 being "always" an average of that is considered as well. Some states have multiple features and pumps listed because there was a tie in the score. For example, features DI, TRN, and GPS all scored the highest for Colorado at 3. Looking at the results, one can see that most states are also interested in SN and TRN, while the pumps that are most used are EAS and BR

In Figure 8, a heat map has been plotted across the United States and Canada to give an understanding of how much people value the features. Most of the values are close to 3.0, with New Mexico at the highest average scored value of 4.64.

Table 2. Variation of features with location

State/ Provinces	Count	Feature(s)	Pump(s)
California	31	SN	EAS
Colorado	8	DI, TRN, GPS	EAS
Florida	24	SN	BR, EAS
Georgia	6	GPS	EAS
Kansas	6	TRN	BR, EAS
Massachusetts	6	AR	BR
Minnesota	7	UPC	BR, EAS
Missouri	7	TRN, SN	BR, EAS
New Jersey	7	TRN, SN	BR
New York	7	TRN	BR
North Carolina	6	TRN, SN	BR
Ohio	13	SN	EAS
Oklahoma	7	GPS	BR
Ontario	11	TRN	BR
Pennsylvania	7	TRN	BR
Texas	9	SN	EAS
Washington	6	SN	BR
Wisconsin	6	TRN	BR

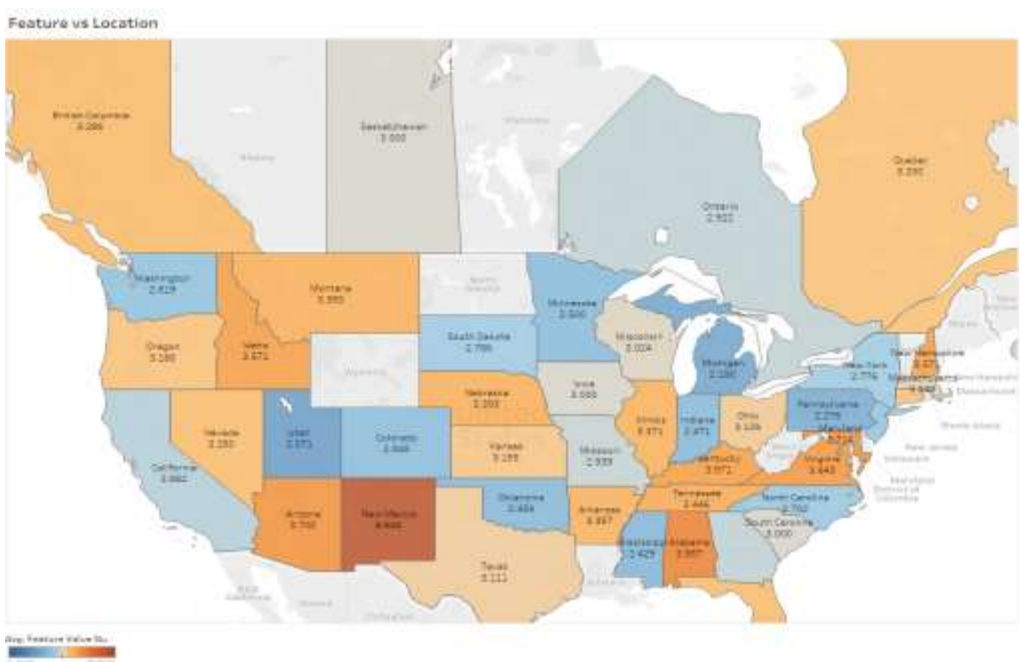

Fig. 8 Variation of all features across location

4.3. Statistical Analysis

The statistical analysis focused on the 14 feature variables from the survey. The analysis is started by examining the preliminary exploration of each variable (Appendix B). No missing variables were noted, as incomplete data points were excluded from the dataset before the analysis was conducted. Two features, Service Notifications (SN) and Tip Replacement Notification (TRN), had a few outliers each. However, since the dataset included over 200 cases and since only one dimension is being considered here, the features are still included in the analysis. There is no significant skewness in the variables in terms of non-linearity.

The individual features were then examined to ensure there was enough level of correlation present to apply interdependence techniques (Appendix C). As all Correlation values were between 0.30 and 0.80, there were no concerns about redundant variables.

The Partial Correlation values for the variables were also within acceptable ranges. Finally, outliers across all the variables were examined through the Mahalanobis distance plot. It was determined that there were no major outliers to exclude.

As the survey was designed with industry experts, there can be a good correlation among variables. Rather than simply reducing correlated variables to a smaller set of important independent composite variables, it is important to test a theoretical model of latent factors causing observed variables. Initially, PCA was conducted (not presented here) and the correlation was found to be greater than 0.32 (Table 3), which motivated the usage of Exploratory Factor Analysis (EFA).

Once the data has been cleaned an initial EFA has been run to determine the number of factors. After examining the Eigenvalues and Scree Plot for the data through Figure 10, two factors have been derived.

Table 3. Correlations between factors

	Correlations	
	Software Features - FBS	Security Features - FBS
Software Features FBS	1.000	0.492
Security Features FBS	0.492	1.000

Eigenvalues

Number	Eigenvalue	Percent	Cum Percent
1	7.5965	54.261	54.261
2	1.2775	9.125	63.386
3	0.7236	5.169	68.555
4	0.6163	4.402	72.957
5	0.5840	4.172	77.129
6	0.5246	3.747	80.876
7	0.4531	3.236	84.112
8	0.4074	2.910	87.023
9	0.3722	2.659	89.682
10	0.3352	2.394	92.076
11	0.3278	2.341	94.417
12	0.2903	2.073	96.490
13	0.2745	1.961	98.451
14	0.2168	1.549	100.000

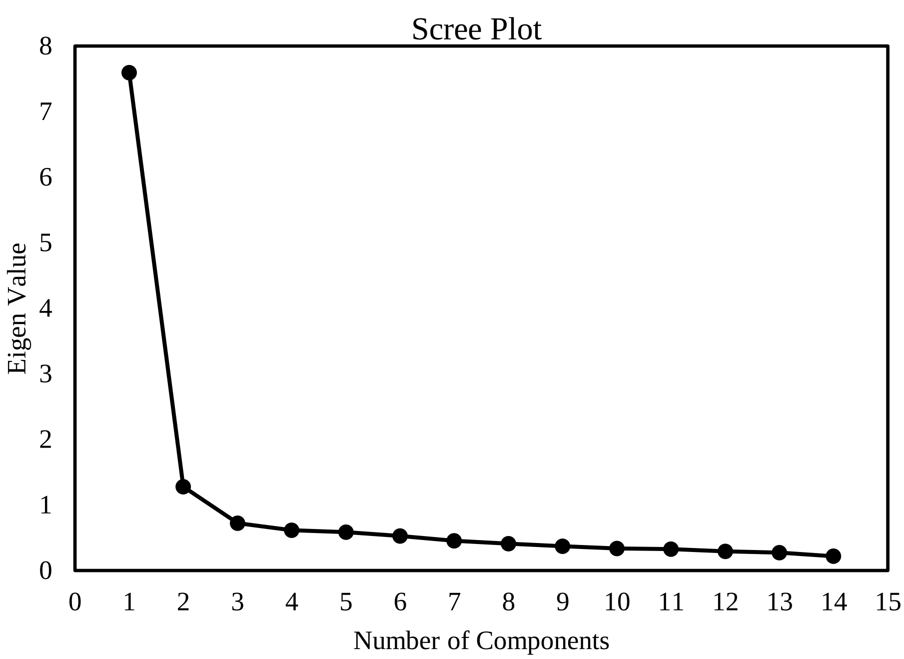

Fig. 10 Eigen values and scree plot

Table 4. Factor loading after the final iteration

Rotated Factor Loading		
	Factor 1	**Factor 2**
RS	0.7100	0.1313
POV	0.8281	-0.0105
GPS	-0.0107	0.7960
RL	0.0923	0.7197
EC	0.6222	0.1673
SPC	0.9348	-0.1504
UPC	0.8166	-0.0378
TSR	0.8806	-0.1198
TRN	0.7134	0.0180
RD	0.6987	0.1207
RP	0.5892	0.1870

No cross-loading is present, and all values are greater than 0.50. This is a good factor to break down. Next, we looked at the Communalities (Table 5).

Ideally, communalities should be greater than 0.50. Even though the variable RP has a value of less than 0.50, it is very close to the margin. It has a decent factor loading, which resulted in including the variable for further analysis.

Table 4. Final communalities for the final iteration

Final Community Estimates	
RS	0.6188
POV	0.6767
GPS	0.6249
RL	0.5958
EC	0.5240
SPC	0.7495
UPC	0.6359
TSR	0.6796
TRN	0.5227
RD	0.5909
RP	0.4973

Lastly, the variance from each factor is studied (Table 5). While 60% is required to trust a factor analysis model, here it is observed that 79% of the variance in the data can be explained by the two factors that were derived. This validates that the statistical model presented here is a good fit for the dataset in consideration.

Table 5. Variance of each factor

Variance by Each Factor			
Factor	**Variance**	**Percent**	**Cumulative Percent**
Factor 1	5.782	52.564	52.564
Factor 2	2.903	26.391	78.955

The next step in the analysis is to develop factor-based scores and check the reliability of each factor (Appendix D). No changes are needed after evaluating the reliability of the factors. Naming the factors required a conceptual understanding of each variable, and it has been determined that Factor 1 is Software Capabilities and Factor 2 as Security Features.

Now that there is a model to employ for exploring the data, a few use cases (hypothesis tests) are studied at a significance level of $\alpha = 0.05$

5. Conclusion

Descriptive and statistical analysis has been conducted on the dataset related to a market survey of smart spray solutions. It is observed that service notifications, tip replacement notifications, and tip service notifications are the major features, the participants showed interest in. Further, an interdependent statistical technique (Factor Analysis) has been employed on the 14 features to reduce the correlated variables into two factors – Software Capabilities and Security Features. With a significance level of 5%, there has been no significant relationship between the factors and several user profiles and demographic variables. However, the experience level of the user has a high statistical significance with software capabilities. Contrary to popular belief that the age group would have significance to the features in this industry, from the present dataset, it has been observed that experience level has more significance.

The next steps in the analysis would be to conduct several other independent statistical techniques like repeated measures ANOVA, chi-square analysis, and regression to study the features individually. This would be the next step in the data analysis and would pave a path for a business analysis team at a paint spraying company, to decide the feature to invest the financial and personnel resources.

References

[1] Yuanyuan Song et al., "Technology Application of Smart Spray in Agriculture: A Review," *Intelligent Automation & Soft Computing*, vol. 21, no. 3, pp. 319-333, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[2] Sinan Akdan et al., "An Agricultural Unmanned Ground Vehicle for Localized Spraying of Harmful Plants Using CNN," *Interdisciplinary Studies on Contemporary Research Practices in Engineering in the 21st Century-III*, pp. 251-265, 2023. [CrossRef] [Publisher Link]

[3] Ramesh K. Sahni et al., *Drone Spraying System for Efficient Agrochemical Application in Precision Agriculture*, Applications of Computer Vision and Drone Technology in Agriculture 4.0, Springer, Singapore, pp. 225-244, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Dan Miller, Smart Sprayer Technology, 2022. [Online]. Available: https://www.dtnpf.com/agriculture/web/ag/news/article/2022/02/10/smarter-sprayers

[5] Paul Kline, *An Easy Guide to Factor Analysis*, 1st ed., Routledge, pp. 1-208, 1994. [CrossRef] [Google Scholar] [Publisher Link]
[6] Jae-On Kim, and Charles W. Mueller, *Factor Analysis Statistical Methods and Practical Issues*, Sage Publications, pp. 1-88, 1978 [Google Scholar] [Publisher Link]
[7] Raymond B. Cattell, "The Scree Test for the Number of Factors," *Multivariate Behavioral Research*, vol. 1, no. 2, pp. 245-276, 1966. [CrossRef] [Google Scholar] [Publisher Link]
[8] Raymond B. Cattell, and S. Vogelmann, "A Comprehensive Trial of the Scree and KG Criteria for Determining the Number of Factors," *Multivariate Behavioral Research*, vol. 12, no. 3, pp. 289-325, 1977. [CrossRef] [Google Scholar] [Publisher Link]