TO WHOMSOEVER IT MAY CONCERN

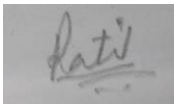
This is to certify that Ms **Ashwati Shrimali** has successfully completed the internship programme of 3 to 4 weeks starting from **20/06/2021** to **14/07/2021** at **D.Y. Patil International University** under the guidance of **Ms. Priti Patil.**

We wish her best future endeavours.

Signature

Candidate

Signature



Faculty Mentor

Signature
Internship Coordinator

School of Computer Science, Engineering and Applications, D Y Patil International University, Akurdi, Pune 411044, Maharashtra

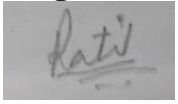
STUDENT INTERNSHIP PROGRAM APPLICATION & ACCEPTANCE

Complete and submit to the TPO/ Internship Program Coordinator. Type or write clearly.

1. Student Name:	Ashwati Shrimali	
2. Campus Address:	Sector 29,Nigdi Pradhikaran,Akurdi,Pune, Maharashtra 41104	Phone:+91-9071123434
3. Home Address:	D/23 Navkar Appt, Ahmedabad	Phone:7203841170
3a. Student email address: 20190802104@dypiu.ac.in		
4. Academic Concentration: AI-ML	5. Internship Semester:2nd Year.	
6. Overall GPA: Yet to receive		
9. Internship Preferences: NA		

	Location	Core Area	Company/ institution
Preferance-1	NA	NA	NA
Preferance-2	NA	NA	NA
Preferance-3	NA	NA	NA

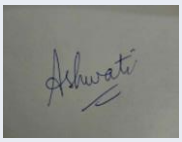
Faculty mentor Signature:



Date: 14/08/2021

Signature confirms that the student has attended the internship orientation and has met all paperwork and process requirements to participate in the internship program,and has received approval from his/her Advisor.

Student Signature:



Date: 14/08/2021

Signature confirms that the student agrees to the terms, conditions, and requirements of the Internship Program

ATTENDANCE SHEET

(For 4 years Degree Programme. / M.Tech. & MBA)

Name & Address of Organization

Name of Student	Ashwati Shrimali
Roll. No	20190802104
Name of Course	B.Tech in Computer Science and Engineering
Date of Commencement of Trg.:	20/06/2021
Date of Completion of Training:	14/07/2021

Initials of the student

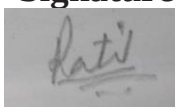
Jun & Jul - 2021	20	21	22	23	24	25	26	27	28	29	30	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS	AS

Note :

1. Attendance Sheet should remain affixed in Daily Training Diary. **Do not remove or tear it off.**

2. Students should sign/initial in the attendance column. Do not mark ‘P

Signature of Mentor



(Name: Ms. Priti Patil) **Contact No.+91 9405015958**

OBJECTIVES/ GUIDELINES/ AGREEMENT: INTERNSHIP SYNOPSIS (THIS WILL BE PREPARED IN CONSULTATION WITH FACULTY MENTOR)

An internship is a unique learning experience that integrates studies with practical work. This agreement is written by the student in consultation with the faculty Mentor and Industrial supervisor. It shall serve to clarify the educational purpose of the internship and to ensure an understanding of the total learning experience among the principal parties involved.

Part I: Contact Information

Student

Name: Ashwati Shrimali Student ID# 20190802104 Class Year: 2nd Year
Campus Address: Sector 29, Nigdi Pradhikaran, Akurdi, Pune, Maharashtra, 411044
City, State: Pune, Maharashtra
Phone: 7203841170 Email: 20190802104@dypiu.ac.in

Faculty Mentor

Name: Ms. Priti Patil Phone: +91 9405015958

Campus Address: Sector 29, Nigdi Pradhikaran, Akurdi, Pune, Maharashtra, 411044

Academic Credit Information

Internship Title: Neuroscience Data Analysis

Department: B.Tech Computer Science

Course #: Computer science and engineering Credits: 4 to 6

Grading Option: Credits Credit/Non-credit: Credit

Beginning Date: 20/06/2021 Ending Date:14/07/2021

Hours per Week: 25 hours Internship is: Unpaid

SUMMER INTERNSHIP REPORT
Neuroscience-Data-Analysis

(20/06/2021 - 14/07/2021)

Prepared by

Ashwati Shrimali - 20190802104
Rutuja Kulkarni - 20190802098
Ritika Kumari - 20190802065
Paras Thakkar - 20200812001
Nidhi Sinha - 20190802066

Faculty Mentor

Ms. Priti Patil

Date of submission

(14/07/2021)
IVth Semester

Declaration

I hereby declare that the report namely “Neuroscience Data Analysis” is completed by us which is based on our practical work experience and a comprehensive study of the existing activities of the Data science department.

We also declare that this report acknowledges the original work by the researcher and specifically attributes our analysis based on the existing dataset. This particular report has not been previously submitted to any other University/College/Organization for academic qualification/ certificate/ diploma or degree.

We have prepared it for the academic purpose of a Bachelor of Technology degree which requires practical work experience.

Ashwati Shrimali
Rutuja Kulkarni
Ritika Kumari
Paras Thakkar
Nidhi Sinha

Summer Internship
Project Approval & Certification

Department of Computer Science and Engineering

D Y Patil International University

AKURDI, PUNE

The project entitled Neuroscience Data Analytic for educational data submitted by Ms. Nidhi Sinha, Ms. Rutuja Kulkarni, Ms. Ritika Kumari, Ms. Ashwati Shrimali, Mr. Paras Thakkar is approved for Summer Internship 2021 programme from 20th June 2021 to 10th July 2021, at Department of Computer Science and Engineering, D.Y. Patil International University.

Acknowledgments

This dataset is from the research article "High-accuracy detection of early Parkinson's Disease (henceforth called PD) using multiple characteristics of finger movement while typing" by Warwick R. Adams. We are trying to analyse the relationship between a few core parameters pertaining to keystroke data viz. HoldTime, LatencyTime & FlightTime. Also how those parameters are related to the Neurological Condition (Parkinson's Disease) of the Users.

We would like to thank my academic supervisor Ms. Priti Patil, Mentor, D Y Patil International University for providing us all the necessary help for the completion of this report. Thank you very much ma'am for guiding us to start and complete this report successfully. We are also thankful for your patience that you have shown during the project.

We would like to express our special thanks to our mentor Ms. Priti Patil Ma'am, who gave us the golden opportunity to do this internship Neuroscience-Data-Analysis, who also helped us in completing our internship. We came to know about so many new things. We are really thankful to them.

Secondly we would also like to thank all who helped us a lot in finalizing this project within the limited time frame.

Executive Summary

Our attempt is to study the real-world dataset collected by the researcher/author with the help of a few subjects (users) who have a medical condition called Parkinson's Disease. Movement disorder is one of the manifestations of this Neurological Disorder which affects the operation of keystrokes on the computer keyboard.

Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. In today's business world, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively.

In this project we analysed existing neuroscience datasets to find some pattern, impacts of certain factors on neurological conditions, etc. There can be multiple dimensions to it.

For data inspection we took out various information about the data file. In data cleansing we cleaned the garbage data and extracted outliers. In data visualization we plotted the graphs, took out correlation and linear regression.

Table of Contents

List of Figures

List of Tables

CHAPTER - 1

1.1 Introduction -

The problem

Parkinson's Disease (PD) is a progressive neurodegenerative movement disease affecting approximately 2% of people at the age of 65 and is the second most commonly occurring neurodegenerative disease in the elderly (after Alzheimer's Disease), with more than 6.3 million people worldwide with PD. In PD sufferers, loss of dopamine-producing neurons results in a range of both motor and non-motor symptoms and currently there is no cure, no means of slowing the disease progression, and no means of prevention. From the perspective of patient quality of life, PD is one of the most severe of all chronic diseases.

At present, diagnosis relies on observation of a combination of visible symptoms by a specialist (typically a neurologist), however PD is commonly either misdiagnosed or the diagnosis is missed completely. Pagan found that, based on a UK autopsy study, there was a misdiagnosis rate of 24% and that this strongly depended on who was performing the diagnosis and whether or not they were applying diagnostic criteria based on clinical guidelines. Specialists who were not movement disorder experts had a correct diagnosis rate of only 75% and diagnoses by primary care doctors had a correct diagnosis of just 53%. In contrast, movement disorder specialists were mistaken by only 6% to 8%, which raises an obvious issue–in order to be referred on to a movement specialist, the patient's primary health practitioner must first recognise and diagnose the symptoms. In addition, a patient may have the disease for 5 to 10 years before it is diagnosed and, by the time of diagnosis, typically 70% of the neurons in the affected part of the brain (the substantia nigra) have already been lost.

With regard to disease diagnoses more generally, there are various physical and functional biomarkers which can be used to provide both diagnostic and predictive information (for example in predicting responses to therapies and drugs). Human–computer interaction (HCI) researches the interfaces between people and computers, producing markers that can be used to measure the state of the user, for example, physiological, cognitive and mental states. In principle, any device which users interact with, and which produces an output that can be measured and stored, could be utilised as part of HCI—devices such as computers, smartphones, tablet computers, gaming platforms and wearable devices. However, technology-based assessments must also provide valid and

accurate results, be independent of rater's training, and allow easy and repetitive use.

PD results in a range of both motor and non-motor symptoms, with the effects on movement including slowness, sidedness, jerkiness and tremors. The hypothesis of this research study was that PD could be detected in its early stages in a person by changes in the characteristics of finger movement as they typed on a keyboard, and that such changes could be used to distinguish and classify people with PD from those without the disease.

There have been previous HCI studies into the diagnosis of PD using features of movement and gait, speech analysis, gripping and lifting tasks, finger tapping tests, hand and finger movement, and handwriting. To date, such studies have all shown limitations in one or more facets—in the sensitivity and specificity of results, the level of specialist supervision or intervention needed, or the requirement for specialised equipment—which have prevented their application more generally as tools to detect or diagnose PD.

1.2 Significance of the Study

Generally, by the time of diagnosis of PD, the disease is already well advanced, significant neuron loss and damage has already occurred, and any possibility of delaying further disease progression or providing neuroprotection is unlikely. The goal must be to diagnose and treat PD well before the irreversible destructive changes have taken place, ideally at least 5 years earlier than is currently the case. In addition, because the most severe symptoms occur in the advanced stages of the disease, strategies aimed at early detection and treatment will have the most benefit.

The objective of this study is to find the relation between different keystroke timeline patterns viz. HoldTime, LatencyTime & FlightTime

1.3 HoldTime

The elapsed time (ms) between the Key Down and Key Up events when pressing and releasing a key. Hold Times are typically in the range 60 to 140 ms.

1.4 Latency

The elapsed time (ms) from the Key Down of one key until the Key Down of the subsequent Key. Latency between successive keystrokes can be separated into the same-hand (LL and RR) and the opposite-hand (LR and RL). Since the

space key can be pressed by either hand, it was treated separately (LS,RS, SL, SR(. Latency is typically anywhere in the range of 50 to 800 ms (anything greater than that was considered to be a pause in typing).

1.5 Flight Time

The elapsed time (ms) between releasing a key and pressing the subsequent key i.e., Latency – HoldTime

Keystroke Data	Definition	Comments
Timestamp	The time of day (hh:mm:ss.sss).	The time at which each keystroke began.
Hold Time	The elapsed time (ms) between the Key Down and Key Up events when pressing and releasing a key.	Hold Time is a measure of how quickly the finger is tapped and can also indicate the relative force of tapping. Hold Times are typically in the range of 60 to 140 ms.
Latency	The elapsed time (ms) from the Key Down of one key until the Key Down of the subsequent key.	Latency between successive keystrokes can be separated into same-hand (LL and RR) and opposite-hand (LR and RL). Since the space key can be pressed by either hand, it was treated separately (LS, RS, SL, SR). Latency is typically anywhere in the range of 50 to 800 ms (anything greater than that was considered to be a pause in typing).
Flight Time	The elapsed time (ms) between releasing a key and pressing the subsequent key.	*Flight Time = Latency − Hold Time*

https://doi.org/10.1371/journal.pone.0188226.t001

Figure 1

CHAPTER - 2

2.1 Symptoms of Parkinson's Disease

The cardinal features of PD are tremor, bradykinesia, postural instability, muscle rigidity and motor blocks, however there are also a wide range of other motor and non-motor symptoms.

Rest tremor is the most common and easily recognised symptom of PD, present in 70% to 75% of cases. The tremors occur at a frequency of 4 to 6 Hz and are prominent at the distal part of an extremity such as the hands and can also involve lips, chin, jaw and legs. Rest tremors typically disappear with action and during sleep.

Bradykinesia is characterised by a slowness of initiating voluntary movement and in sustaining repetitive movements with progressive reduction in speed and amplitude and is the most characteristic feature of PD. Bradykinesia is symptomatic of all basal ganglia disorders and is typified by difficulty with performing sequential and simultaneous tasks. According to Jahanshahi et al. the initial manifestation of PD is often slowness in performing the normal activities of daily life, especially those tasks requiring fine motor control.

Particularly in its early stages, PD is characterised by a predominantly unilateral (asymmetrical) appearance of the motor symptoms. This sidedness can be so conspicuous that it often serves as a clinical parameter to differentiate the disease from other neurodegenerative Parkinsonian syndromes, for example, in multiple system atrophy, diffuse Lewy body disease and progressive supra-nuclear palsy, there is usually no side predominance.

Rigidity in PD sufferers is characterised by increased resistance, present throughout the range of movement of a limb. When accompanied by an underlying tremor it results in a cogwheel phenomenon, which continues throughout an entire range of movement. The rigidity may occur at the neck, shoulders or hips (proximally) or wrists and ankles (distally).

2.2 Issues with early diagnosis

The problems with diagnosing PD arise because there is no definitive test, and currently the disease diagnosis must be based on clinical and observational

criteria only. Many of the symptoms of PD are imprecise and also common to other diseases, both neurodegenerative and non-neurodegenerative in nature. Evaluation may be performed using the Unified Parkinson's Disease Rating Scale (UPDRS), a tool based on a score derived from the neurological evaluation that is performed by a physician, and hence it is a subjective measure which leads to a lack of objectivity, repeatability and sensitivity in the scale.

Parkinson's disease is usually preceded by a premotor phase that can last for years, or even decades, between the onset of neurodegeneration and manifestation of the classic clinical motor symptoms. The most common pre-diagnostic symptom of Parkinson's Disease within 2 years before diagnosis is tremor, with 41% of individuals reporting symptoms to their medical practitioner compared with less than 1% of controls, and the incidence of tremor is already higher at 5 and 10 years before diagnosis. Despite the reliance on motor symptoms for the standard diagnosis of PD, premotor symptoms hold promise for the early diagnosis of PD and considerable progress has been made in recent years in establishing premotor symptoms as a means of identifying PD much earlier. Biomarkers hold promise for reliable early PD diagnosis, while neuroimaging and sonography show enormous potential for high degrees of sensitivity and specificity in diagnosing early PD.

2.3 Biometrics and keystroke dynamics

The use of keystrokes as a means of identification has a long history. Das et al. used keystroke dynamics while typing a computer login string to identify users with 90% to 99% accuracy. Their technique involved key hold times and latency, using a Gaussian mixture model and a neural network. This, and many other similar studies, demonstrates that keystroke characteristics can be used very accurately to classify the features of particular users.

Keystroke dynamic features can be extracted using the timing information of the key press and release events. The hold time of individual keys and the latency between keys (the time interval between pressing one key and a succeeding key) are typically exploited. In addition to ordered pairs (two successive keystrokes), n-tuples of a sequence of keystrokes have also been investigated and keystroke biometrics research has utilised many machine learning and classification techniques.

2.4 Existing studies into Parkinson's biomarkers

It is known that many PD biomarkers can be analysed using various forms of human-computer interaction, including movement and gait analysis, speech

analysis, the precision grip and lift test (PLGT), finger tapping tests (FTT), hand and finger movement, and handwriting.

In non-PD subjects, finger tapping frequency declines with advancing age, men tap faster than women, and tapping with the dominant finger is faster than that of the non-dominant finger. The basal ganglia facilitate sequential movement and the sequence of movements, however in PD patients bradykinesia and disturbances of rhythm formation occur which can be assessed by FTT. Using an accelerometer and touch sensor, Yokoe et al. measured 14 parameters of FTT movement, showing clear differences between PD and non-PD patients. By classifying these into both velocity and amplitude parameters and rhythm-related parameters, they found that maximum opening velocity was the most sensitive measure and most closely aligned with the UPDRS FTT score.

People with PD tend to have slower reaction times than non-affected people of similar age and this can be investigated with regard to the sequence of mental steps that occur between the time that a stimulus is presented and the subsequent physical response. Reaction times can be separated into either a simple response (such as pressing one key) or a complex response (pressing a sequence of keys). Low et al. found that, especially when complex responses are required, the reaction times of PD patients were slower, both in the delayed onset of pre-motor processes and the motor responses themselves. Another effect of PD is increasing difficulty in performing sequential and bi-manual movements. Pal et al. suggested that analysis of sequential hand and finger movements may provide for indication of PD, then Giancardo et al. , prompted by the use of keyboard typing characteristics in biometrics, utilised the typing of people on a computer keyboard as a means of observing and potentially quantifying motor impairment such as in PD. Using a time series analysis of keystroke hold times and a support vector machine (SVM) they showed significant differences between PD patients and controls, but a related follow-on study of typing characteristics with a larger group only achieved an accuracy of 78%. Notably, none of these studies appeared to consider the asymmetry of movement (between left and right hands), although such sidedness is a significant feature in early PD.

A conclusion from previous studies is that PD affects multiple aspects of hand and finger movements and that many of these may be detected (both singly and in combination) through changes in the response characteristics as people type a sequence of text on a computer keyboard.

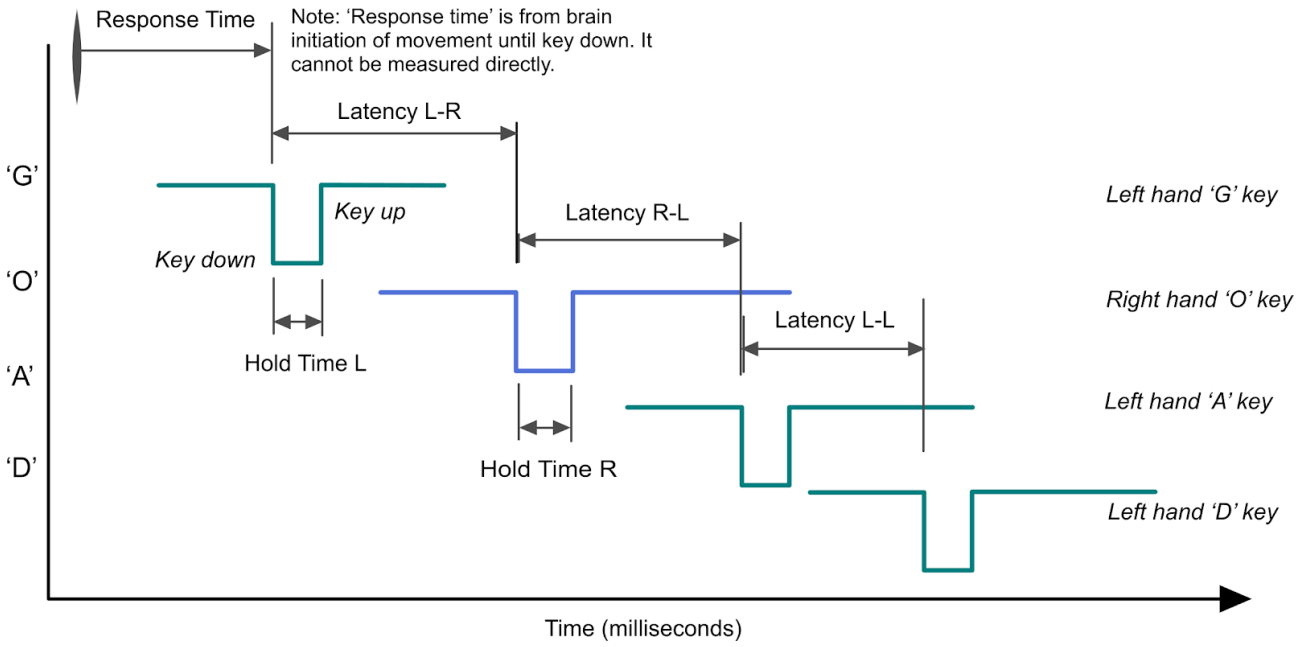


Figure 2

In this investigation, the keystroke dynamics of participants typing on a computer keyboard were captured as they typed normally throughout the day. This was a significant aspect of the investigation, as it meant that the participants were not limited to a directed typing task, the keystroke monitoring was completely un-intrusive upon their normal routine, and no external supervision was involved. The procedure involved each participant installing a small software application ('Tappy'), on their Windows computer, which recorded each keypress event and its timing, along with the key's position on the keyboard and whether it was a left or right-handed key. The specific keystroke timing events that were captured are shown in the table below and the manner in which these timings occur during typing the sequence of letters comprising the word 'Goad' is shown in the figure above. The method involved monitoring all the participants' typing, irrespective of the application they were using at the time (e.g. typing emails and documents), and not being limited to a 'typing test' of predetermined passages of text. This design decision was made in order to facilitate data capture over a longer timeframe, as well as to avoid any stress on the participant to perform well, which could itself change their keystroke dynamics.

CHAPTER - 3

(Our Study/Contribution)

3.1 Objectives

1. Perform data inspection by using describe(), head(), info(), boxplot() etc. functions, data cleansing by removing null values from the dataset if any, and data visualization by creating graphs.
2. Extract outliers from the datasets and duplicates from the csv file.
3. Extract keystroke records (max 100 records) close to threshold value (HoldTime - 60ms to 140ms, LatencyTime - 50ms to 800ms).
4. Find correlations and regression between Hold Time and Latency Time using functions in excel, google sheets and Python.
5. Plot the graph for correlations and regression between Hold Time and Latency Time in google sheets and python.

3.2 Dataset structure description

Each file contains comma separated keystroke data for one month for a particular user. The filename comprises the 10 character code (matching the user details file) and the YYMM of the data. The fields are:

- UserKey: 10 character code for that user
- Date: YYMMDD
- Timestamp: HH:MM:SS.SSS
- Hand: L or R key pressed
- Hold time: Time between press and release for current key mmmm.m milliseconds
- Direction: Previous to current LL, LR, RL, RR (and S for a space key)
- Latency time: Time between pressing the previous key and pressing the current key. Milliseconds
- Flight time: Time between release of previous key and press of current key. Milliseconds

Dataset structure description for users:

The filename of each user file contains a 10 character code, used to cross reference to the keystroke data files for that user. The fields are:

- **Birth Year:** Year of birth
- **Gender:** Male/Female

- **Parkinson’s:** Whether they have Parkinson's Disease [True/False]
- **Tremors:** Whether they have tremors [True/False]
- **Diagnosis Year:** If they have Parkinson's, when was it first diagnosed
- Whether there is sidedness of movement [Left/Right/None] (self - reported)
- **UPDRS:** The UPDRS score (if known) [1 to 5]
- **Impact:** The Parkinson's disease severity or impact on their daily life [Mild/Medium/Severe] (self- reported)
- **Levadopa:** Whether they are using Sinemet and the like [Yes/No]
- **DA:** Whether they are using a dopamine agonist [Yes/No]
- **MAOB:** Whether they are using an MAO-B inhibitor [Yes/No]
- **Other:** Whether they are taking another Parkinson's medication [Yes/No]

3.3 Graphs

1. **User - QAH9IVALVC**

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1942	Femal-e	True	True	2015	Right	Don't know	Medi-um	True	False	False	False

Table 1

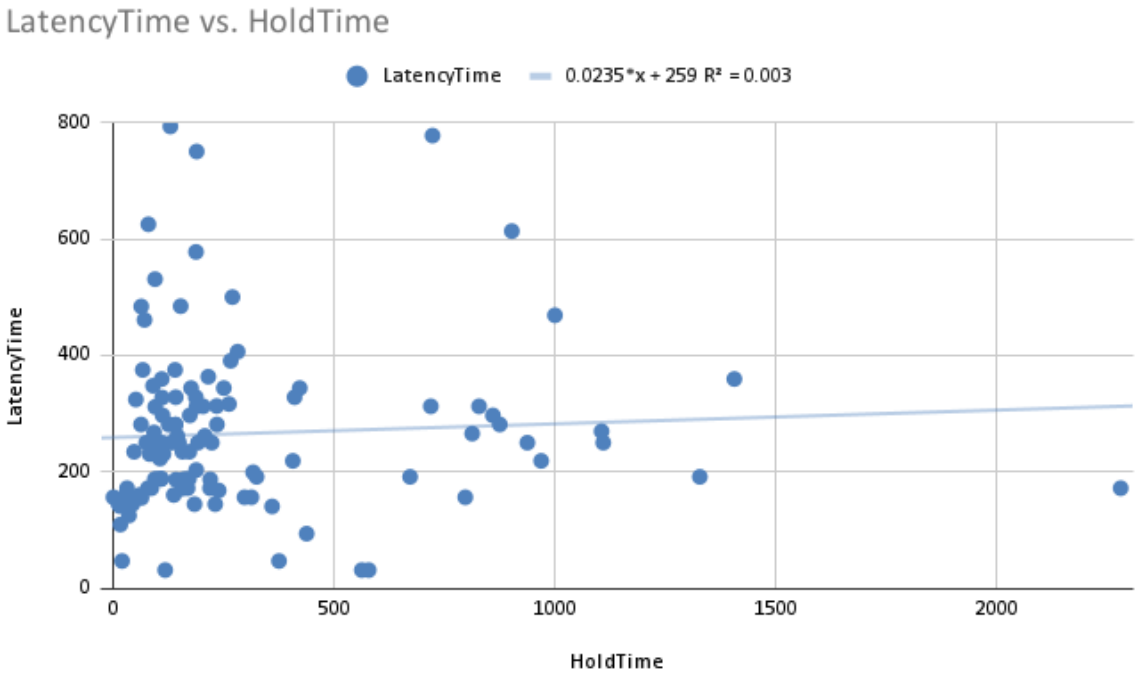


Figure 3

Correlation: 0.05477225575

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.0235 times. LatencyTime and HoldTime are in positive correlation.

2. User - SKLXBAOSN4

Birth Year	Gend-er	Parki nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -B	Other
1951	Femal -e	True	True	2009	Right	Don't know	Medi-um	True	False	False	False

Table 2

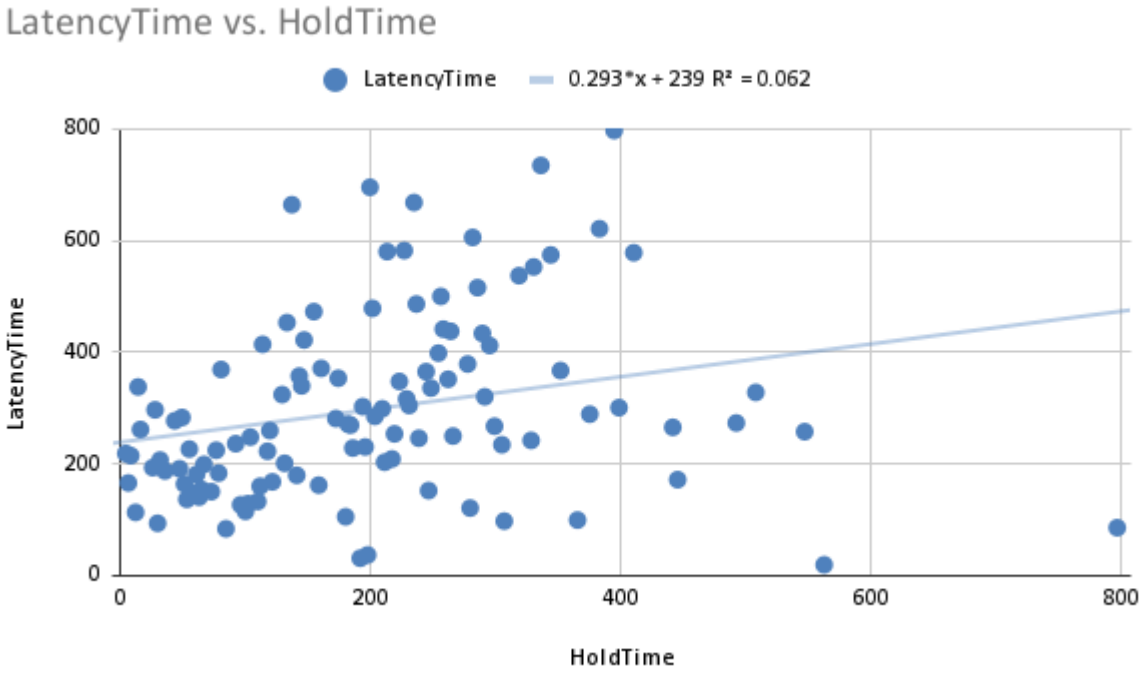


Figure 4

Correlation: 0.248997992

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.293 times. LatencyTime and HoldTime are in positive correlation.

3. User - SPNOI40JCA

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad opa	DA	MAO-B	Other
1947	Male	True	False	2003	Left	3	Medi-um	True	False	False	True

Table 3

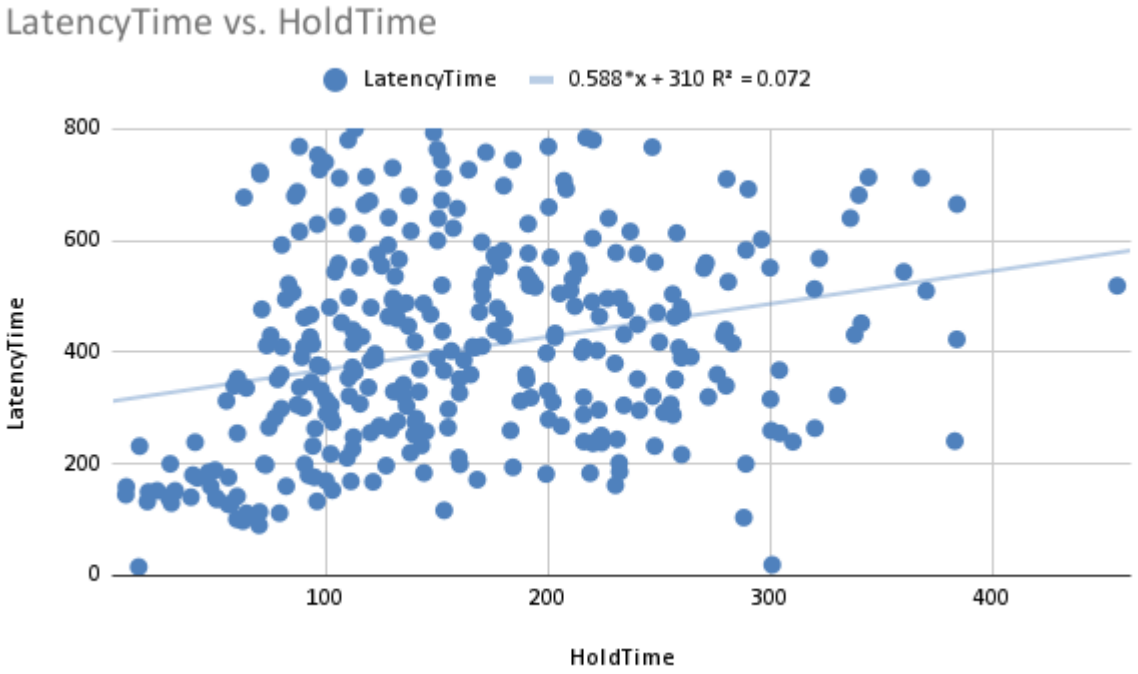


Figure 5

Correlation: 0.2683281573

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.588 times. LatencyTime and HoldTime are in positive correlation.

4. User - TL2XHTLK1T

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1947	Male	True	True	2014	Left	Don't Know	Mild	False	False	False	True

Table 4

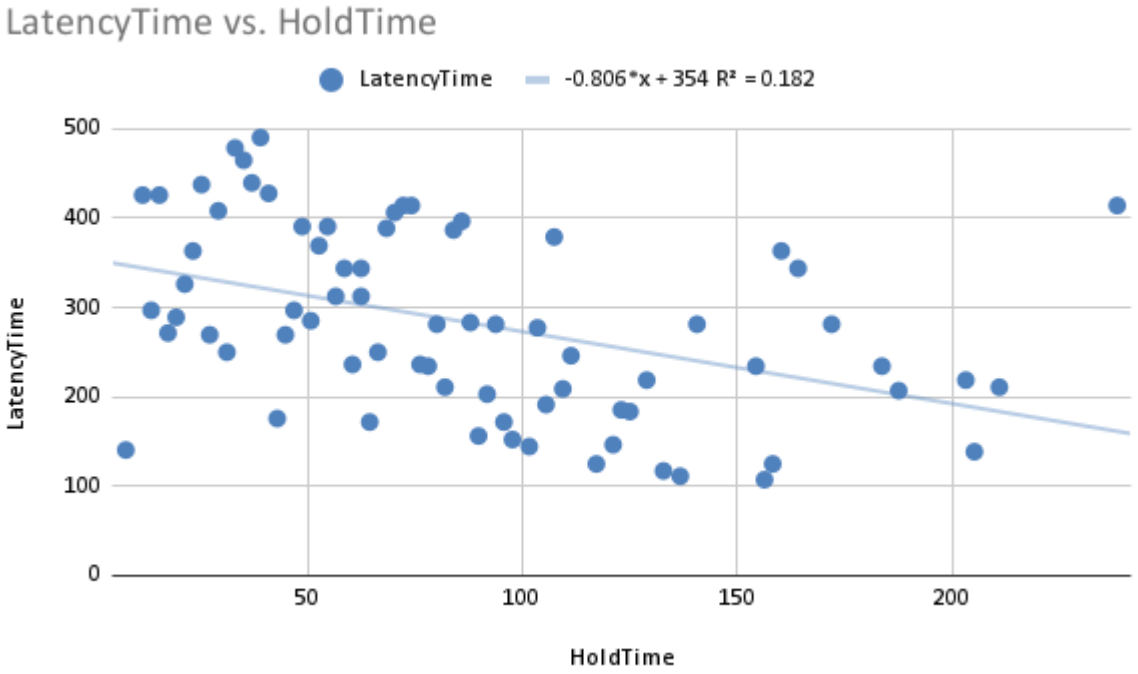


Figure 6

Correlation: -0.4266145802

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to decrease by 0.806 times. LatencyTime and HoldTime are in negative correlation.

5. User - Z2UPVHHGBE

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1947	Femal-e	True	True	2015	Right	Don't Know	Mild	False	False	False	True

Table 5

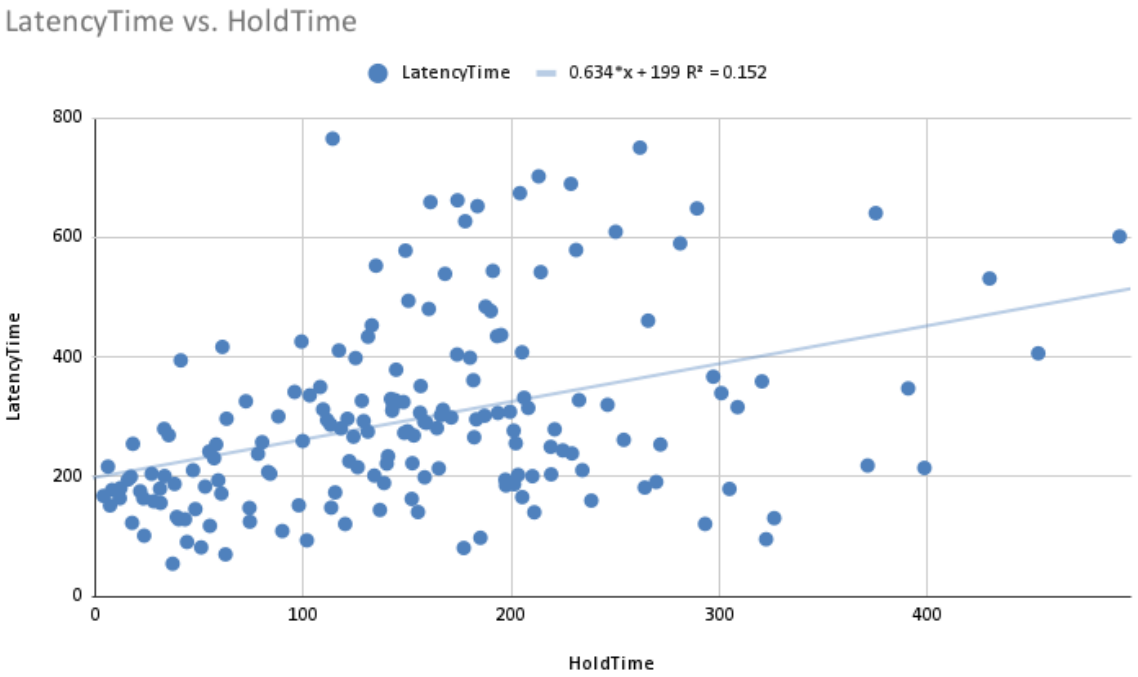


Figure 7

Correlation: 0.3898717738

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.634 times. LatencyTime and HoldTime are in positive correlation.

6. User - ZI1KGKLCD5

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO B	Other
1949	Male	True	False	2015	None	Don't Know	Medi-um	False	False	False	False

Table 6

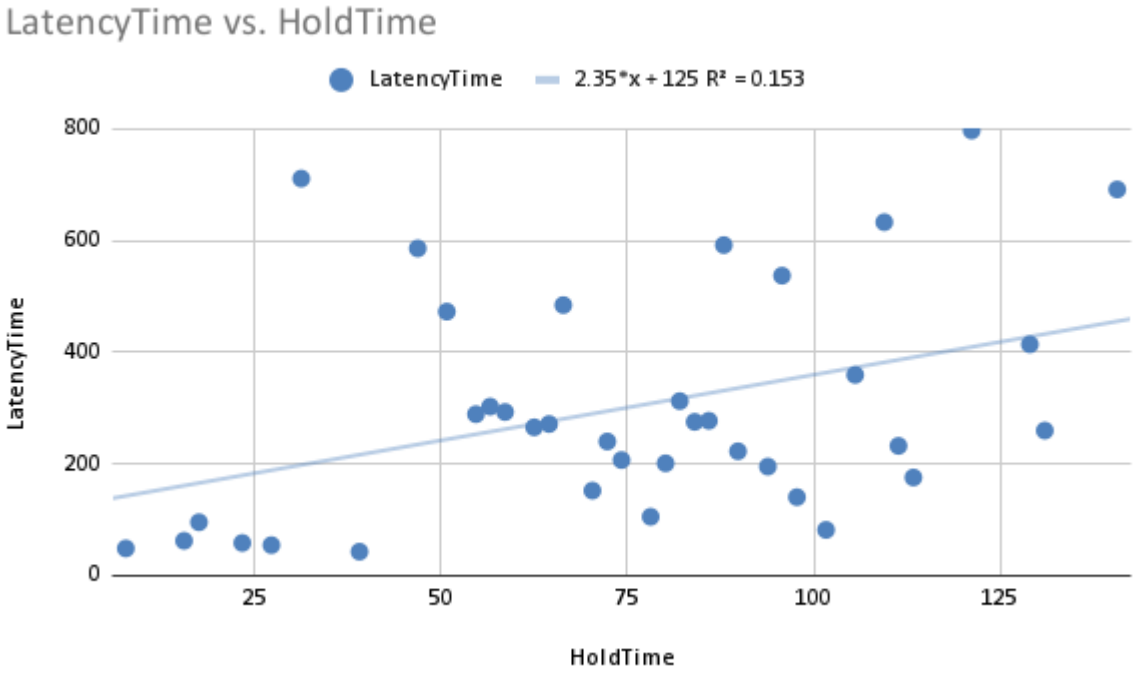


Figure 8

Correlation: 0.3911521443

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 2.35 times. LatencyTime and HoldTime are in positive correlation.

7. User - 48DZPAJ5NS

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -B	Other
1950	Male	True	False	2010	None	Don't Know	Mild	False	False	False	True

Table 7

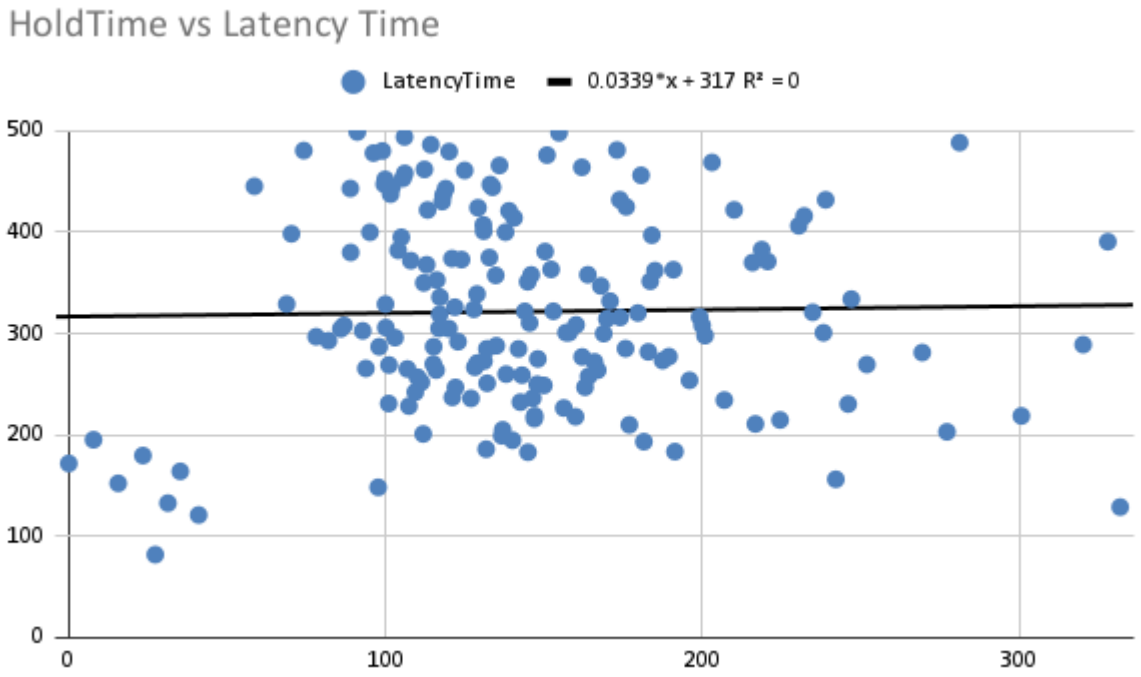


Figure 9

Correlation: 0.020662408

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.0339 times. LatencyTime and HoldTime are in positive correlation

8. User - 0EA27ICBLF

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO B-	Other
1952	Femal-e	True	True	2000	Left	Don't Know	Sever-e	True	True	False	False

Table 8

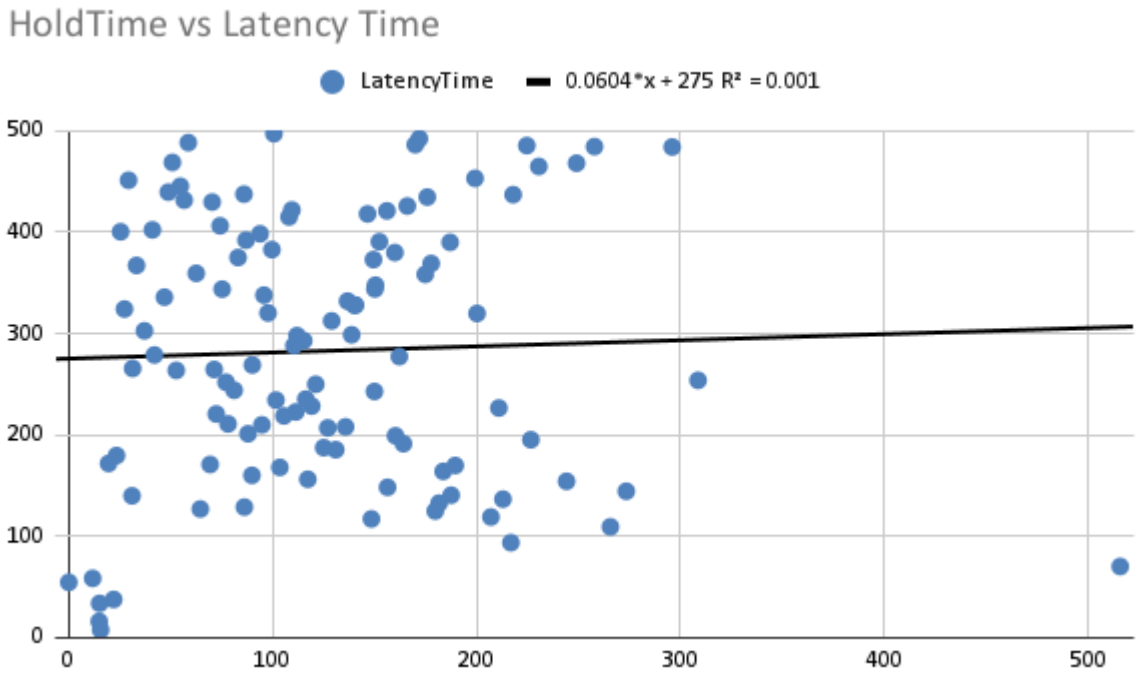


Figure 10

Correlation: 0.03693483

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.0604 times. LatencyTime and HoldTime are in positive correlation.

9. User - 5YFAPLRCMD

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -B	Other
1967	Male	True	True	2016	None	Don't Know	Mild	True	False	False	False

Table 9

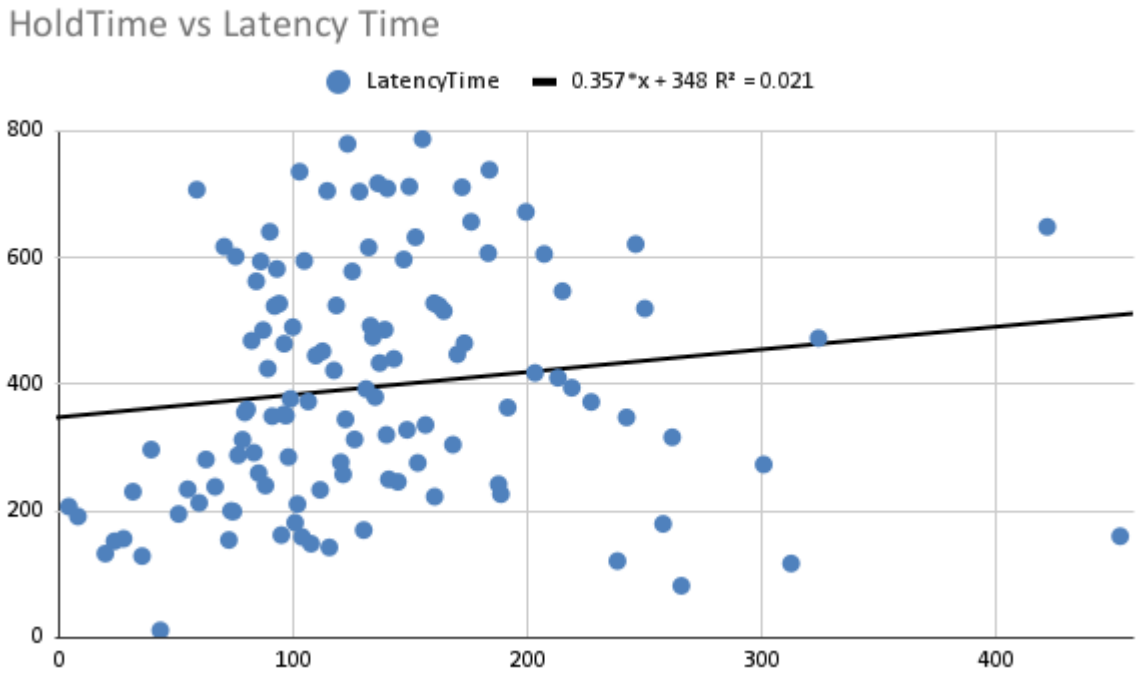


Figure 11

Correlation: 0.143709169

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.357 times. LatencyTime and HoldTime are in positive correlation.

10. User - 5PQVTWULAC

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1943	Femal-e	True	True	2009	Right	Don't Know	Mild	False	True	False	True

Table 10

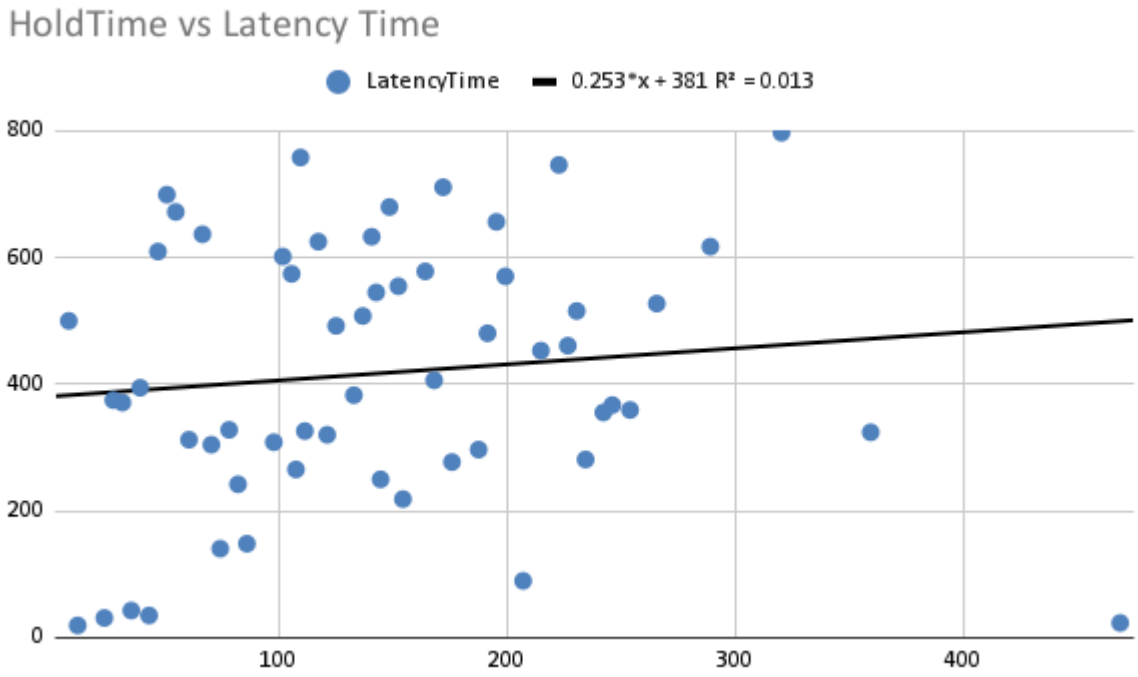


Figure 12

Correlation: 0.113837017

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.253 times. LatencyTime and HoldTime are in positive correlation.

11. User - 2JTCBKUP8T

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1958	Male	True	True	2013	Right	Don't Know	Medi-um	True	False	False	False

Table 11

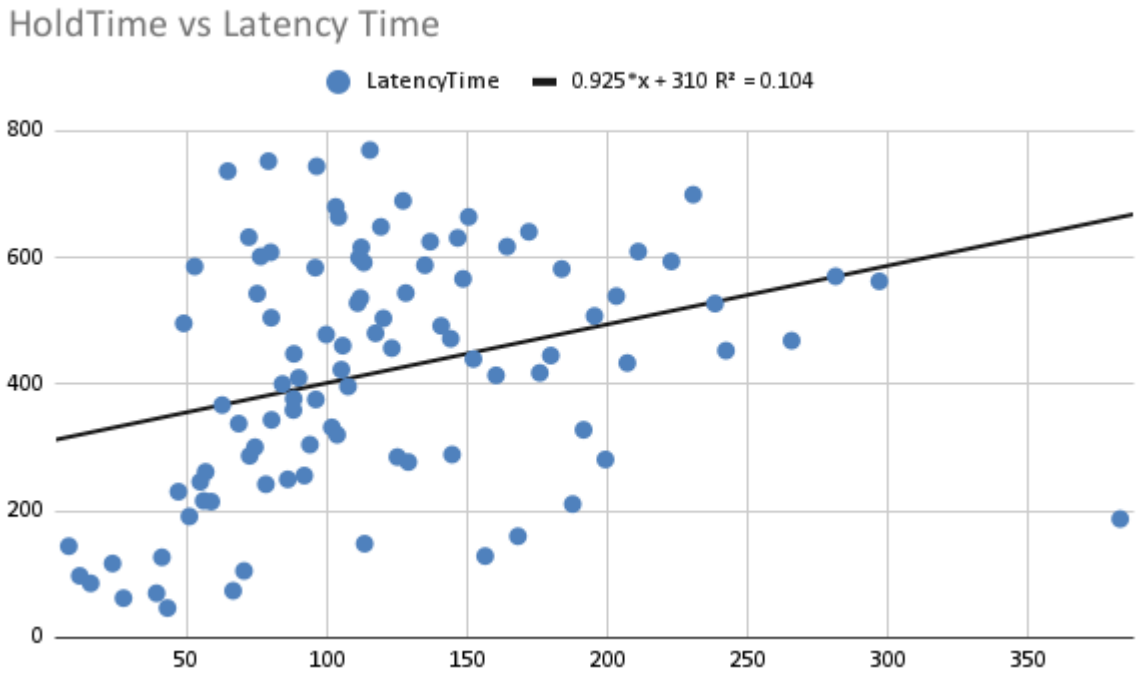


Figure 13

Correlation: 0.322982739

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.925 times. LatencyTime and HoldTime are in positive correlation.

12. User - 5ARV4LHEJ0

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1944	Male	True	False	2016	Right	2	Sever-e	False	False	False	True

Table 12

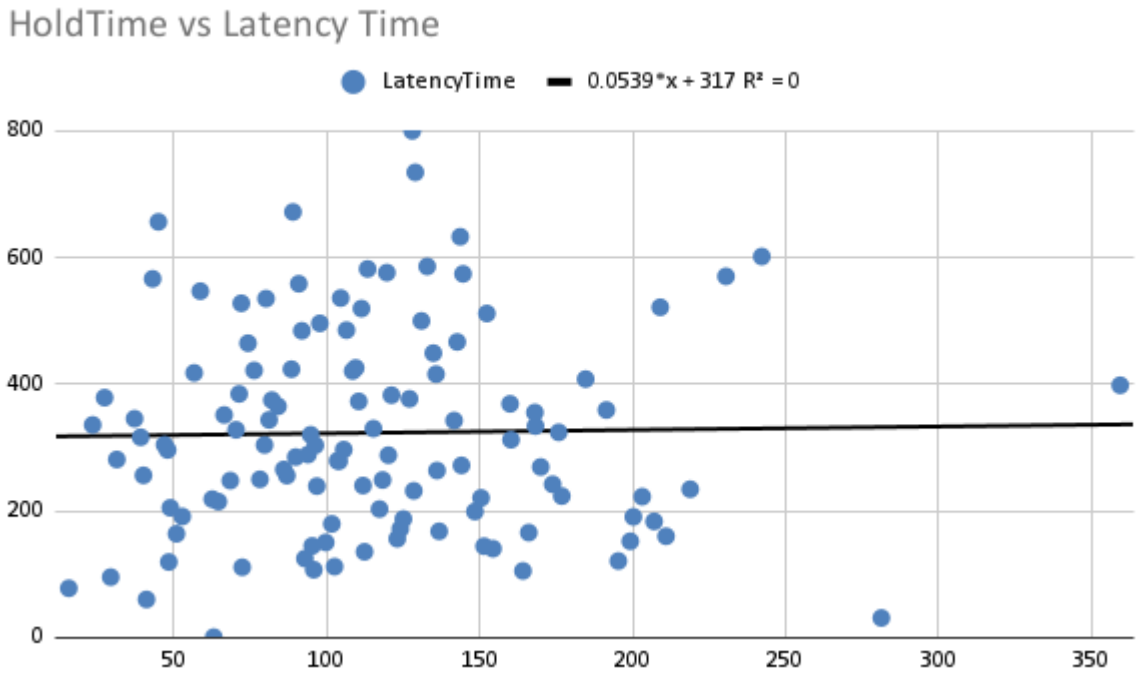


Figure 14

Correlation: 0.018908481

For each Increase of one ms of HoldTime (Independent variable), the expected LatencyTime (Dependent variable) in ms is predicted to increase by 0.0539 times. LatencyTime and HoldTime are in positive correlation.

13. User Key – 95S93AXF86

Birth Year	Gend-er	Parki nson-s	Trem ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -	Other
1952	Femal -e	True	True	2010	None	Don't Know	Medi-um	False	False	False	True

Table 13

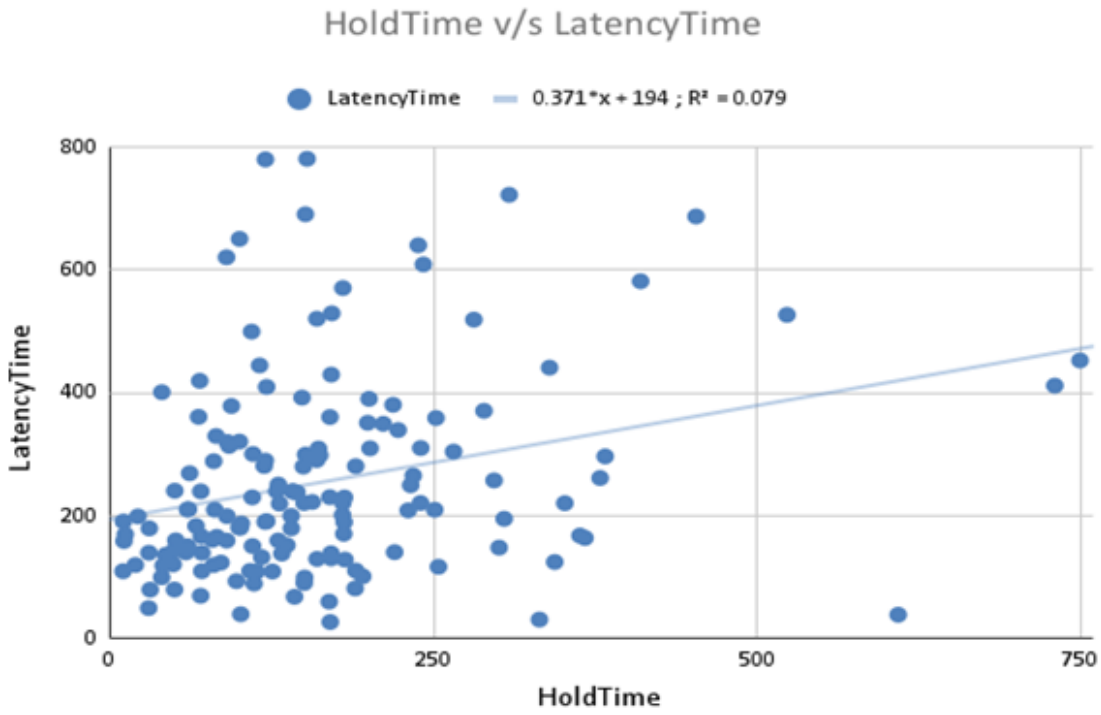


Figure 15

Correlation: 0.281064

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.371 times. LatencyTime and HoldTime are in positive correlation.

14. User Key – BBCWCBNBPR

Birth Year	Gender	Parkinsons	Tremors	Diagnosis-Year	Sided	UPDRS	Impact	Levad-opa	DA	MAO-B	Other
1952	Male	True	False	1996	Left	Don't Know	Medium	True	False	False	True

Table 14

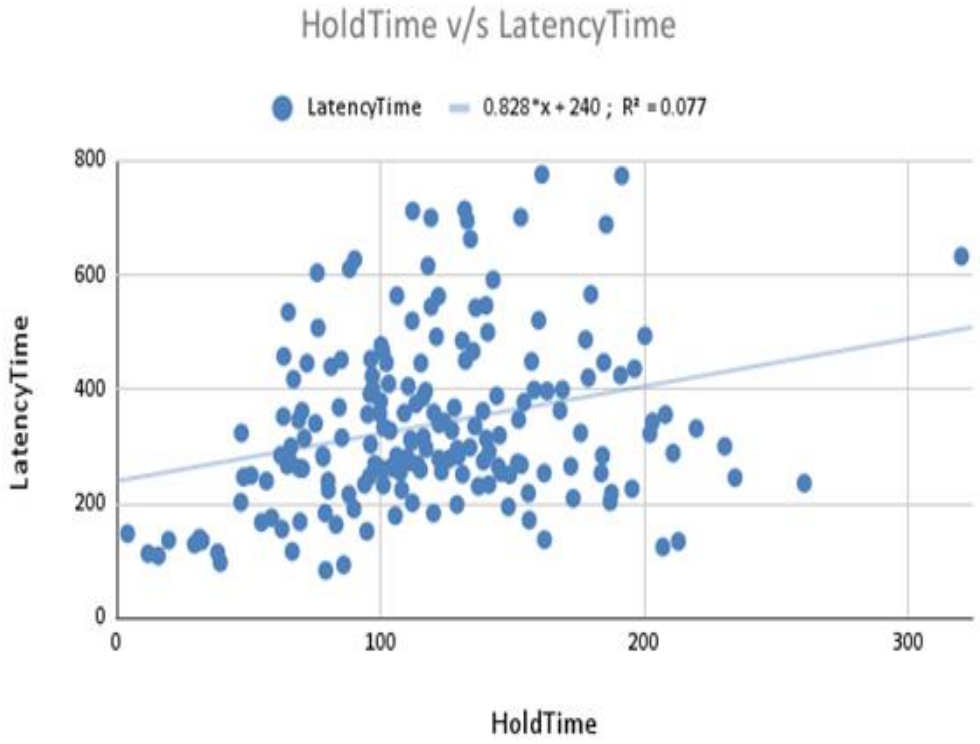


Figure 16

Correlation: 0.27826

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.828 times. LatencyTime and HoldTime are in positive correlation.

15. User Key - CPSM4LIXD

Birth Year	Gend-er	Parki nson-s	Trem ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -B	Other
1946	Male	True	False	2016	Right	Don't Know	Sever -e	False	False	False	True

Table 15

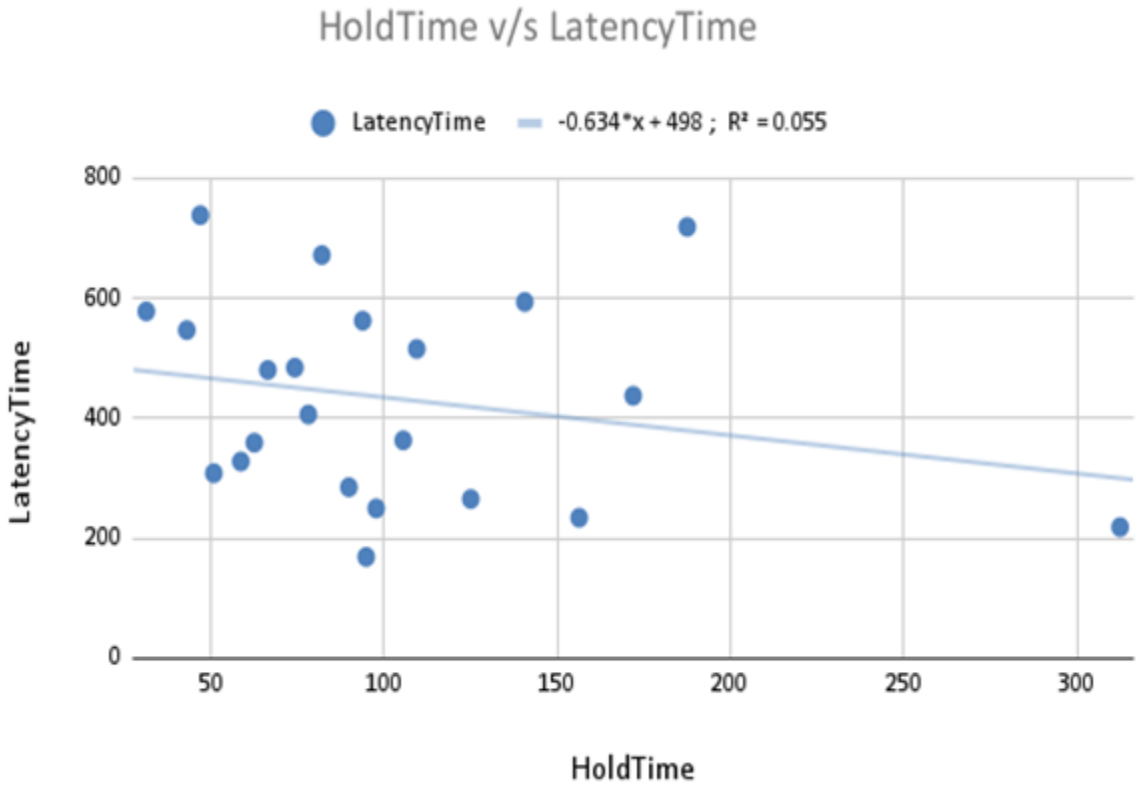


Figure 17

Correlation : -0.23553

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to decrease by 0.634 times. LatencyTime and HoldTime are in negative correlation.

16. User Key – FXC5YFXZ0K

Birth Year	Gend-er	Parki nson-s	Trem ors	Diagn osis-Year	Sided	UPD-RS	Impac -t	Levad -opa	DA	MAO -B	Other
1943	Male	True	False	2009	Left	Don't Know	Medi-um	True	False	False	True

Table 16

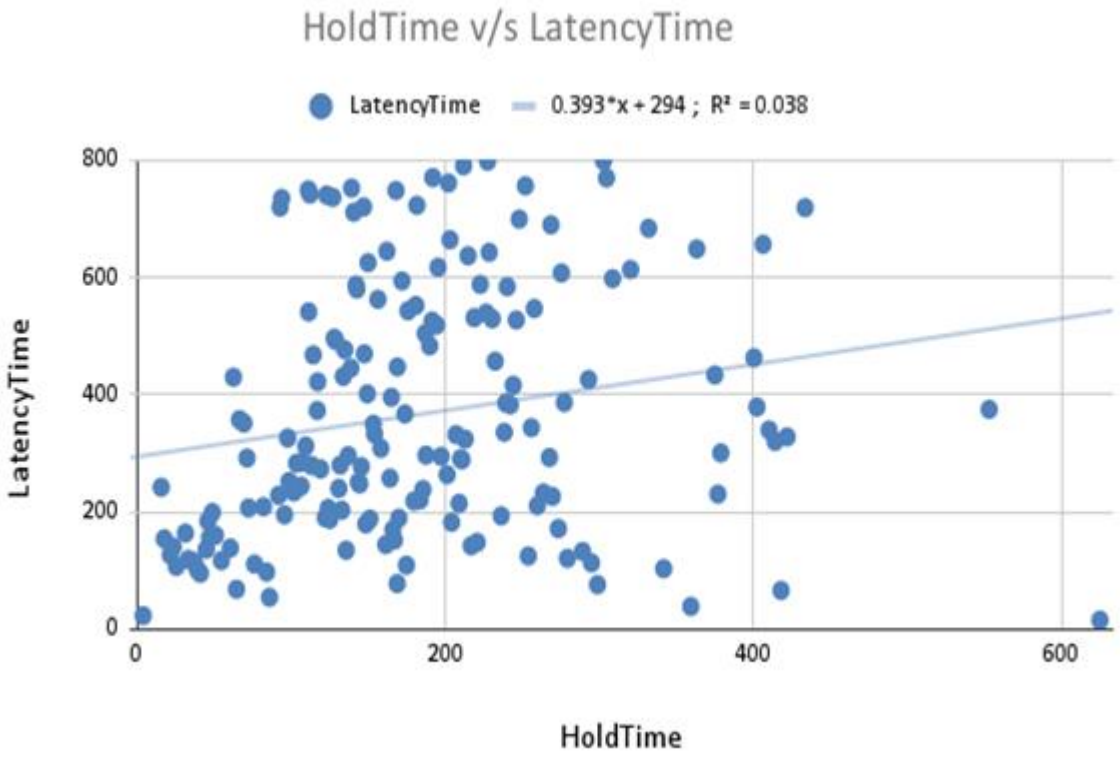


Figure 18

Correlation: 0.194827141

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.393 times. LatencyTime and HoldTime are in positive correlation.

17. User Key – LKYPMARSKU

Birth Year	Gender	Parkinsons	Tremors	Diagnosis-Year	Sided	UPDRS	Impact	Levad-opa	DA	MAO-B	Other
1946	Female	True	True	2017	Right	Don't Know	Mild	False	False	False	False

Table 17

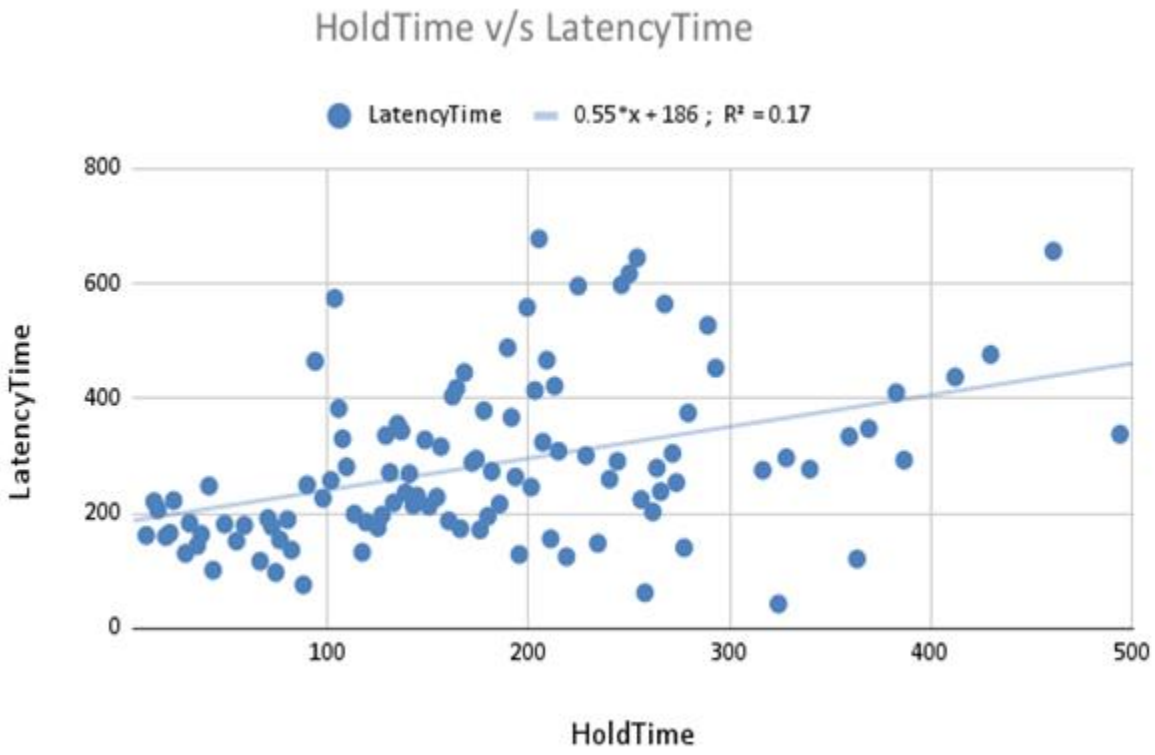


Figure 19

Correlation: 0.412535

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.55 times. LatencyTime and HoldTime are in positive correlation.

18. User Key – LA6KW35OXK

Birth Year	Gend-er	Parki nson-s	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1958	Femal-e	False	True	2016	Left	Don't Know	Mild	False	False	False	False

Table 18

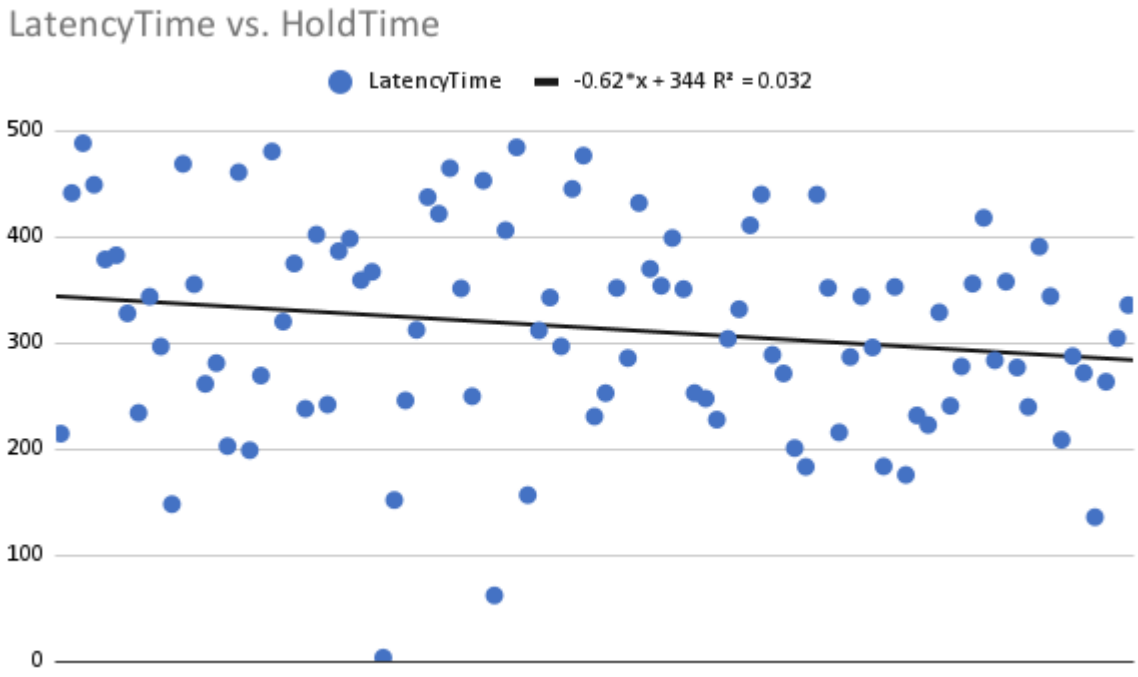


Figure 20

Correlation: 0.366317749

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to decrease by 0.62 times. LatencyTime and HoldTime are in negative correlation.

19. User Key – CYC6IZJYTE

Birth Year	Gend-er	Parki nson-s	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1954	Femal-e	True	True	1998	Left	Don't Know	Mild	True	True	False	False

Table 19

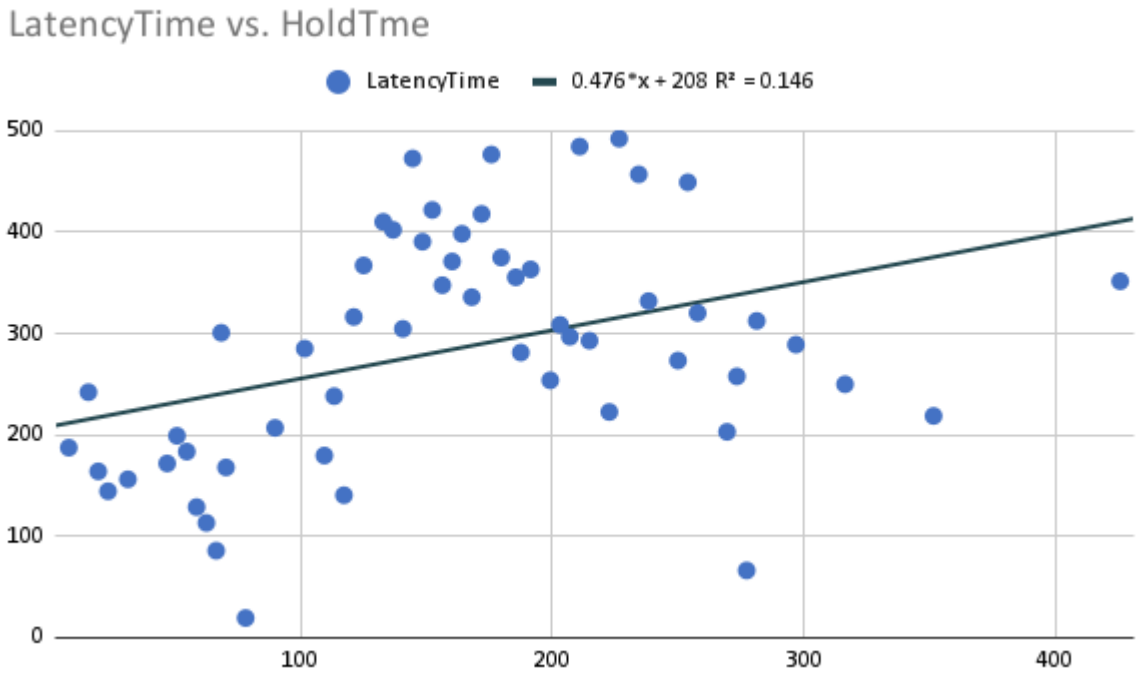


Figure 21

Correlation: 0.382496336

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.476 times. LatencyTime and HoldTime are in positive correlation.

20. User Key –LSQWWDXEYO

Birth Year	Gend-er	Parki-nsons	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1951	Male	True	True	2016	Left	Don't Know	Mild	False	True	False	False

Table 20

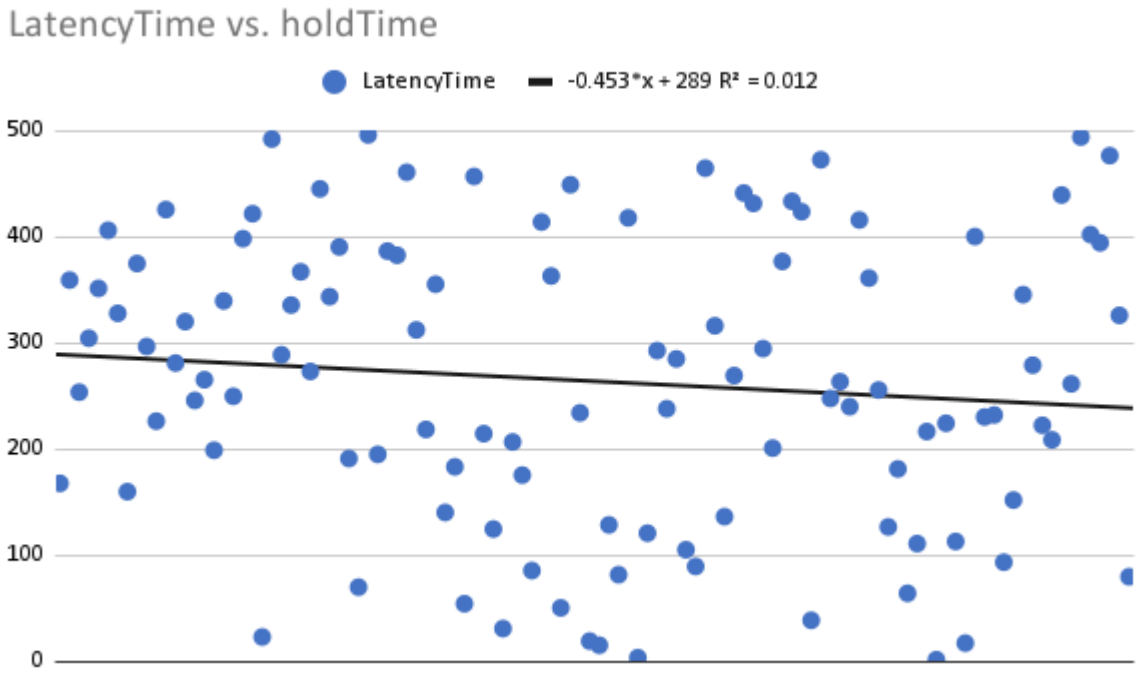


Figure 22

Correlation: -0.061775188

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to decrease by 0.453 times. LatencyTime and HoldTime are in negative correlation.

21. User Key –LVRTXTCS1Z

Birth Year	Gend-er	Parki nson-s	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1932	Femal-e	True	False	--	None	Don't Know	Medi-um	True	False	False	False

Table 21

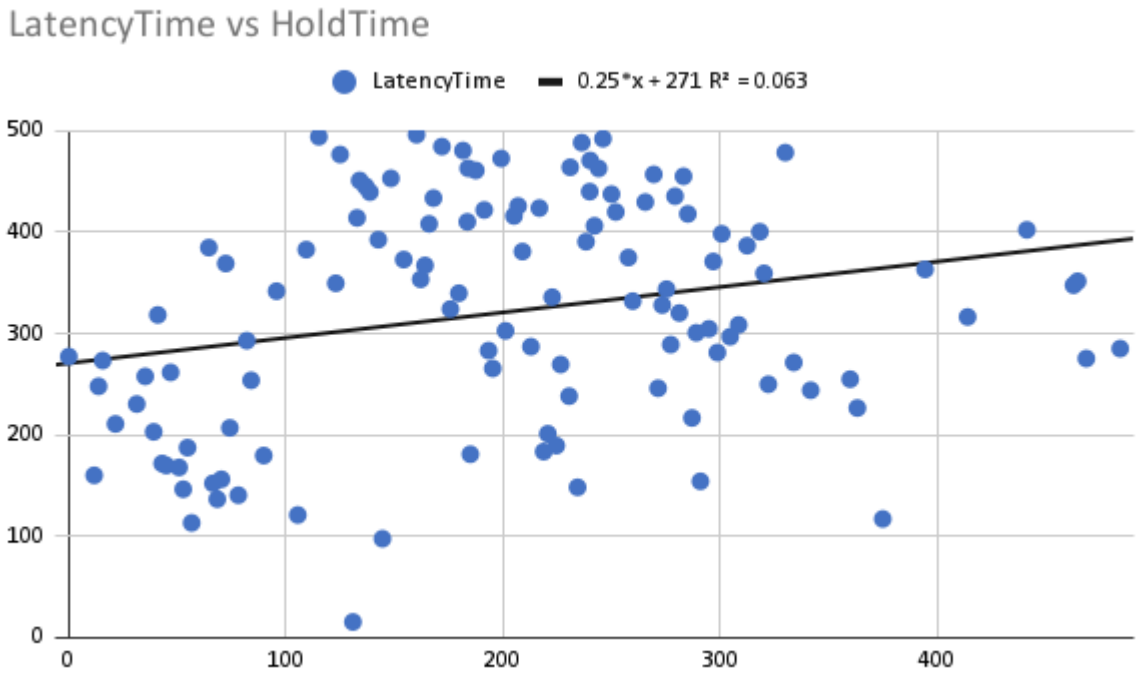


Figure 23

Correlation: 0.2507814891

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.25 times. LatencyTime and HoldTime are in positive correlation.

22. User Key – O0L3RD63QP

Birth Year	Gend-er	Parki nson-s	Trem-ors	Diagn osis-Year	Sided	UPD-RS	Impac-t	Levad-opa	DA	MAO-B	Other
1944	Femal-e	True	True	2012	None	Don't Know	Mild	True	False	False	True

Table 22

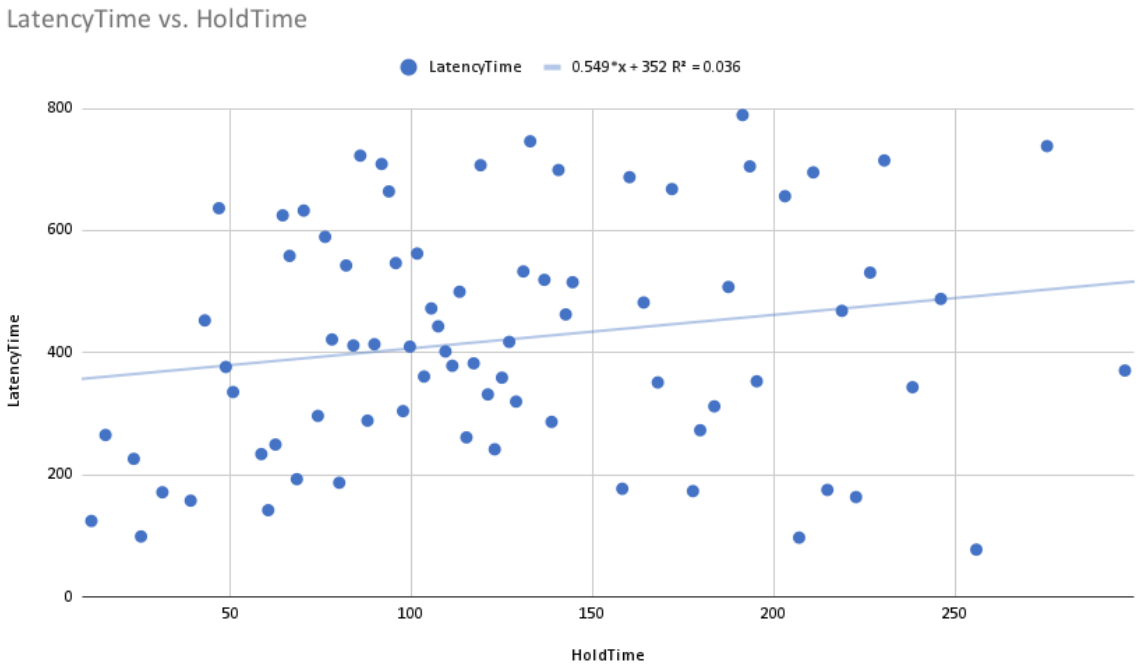


Figure 24

Correlation: 0.1906947153

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.549 times. LatencyTime and HoldTime are in positive correlation.

23. User Key – HB2JCE24IE

BirthY-ear:	Gender:	Parkins-ons:	Tremors:	DiagnosisYear	Sided:	UPDR-S:	Impact:	Levado-pa:	DA:	MAOB:	Other:
1956	Male	TRUE	FALSE	2011	Right	Don't know	Mediu-m	TRUE	FALSE	FALSE	FALSE

Table 23

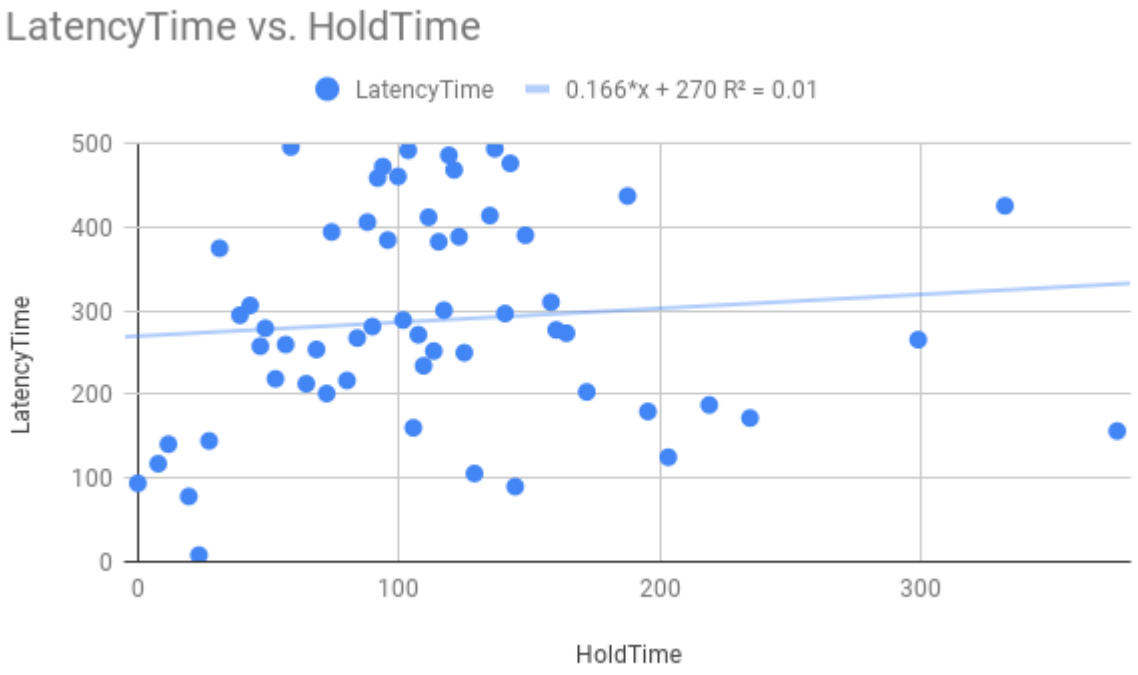


Figure 25

Correlation: 0.09815439251

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.166 times. LatencyTime and HoldTime are in positive correlation.

24. User Key – IDZHIUK2W2

BirthY-ear:	Gender:	Parkins-ons:	Tremor-s:	DiagnosisYear:	Sided:	UPDR-S:	Impact:	Levado-pa:	DA:	MAOB:	Other:
1943	Female	TRUE	TRUE	2013	None	Don't know	Mild	TRUE	False	False	False

Table 24

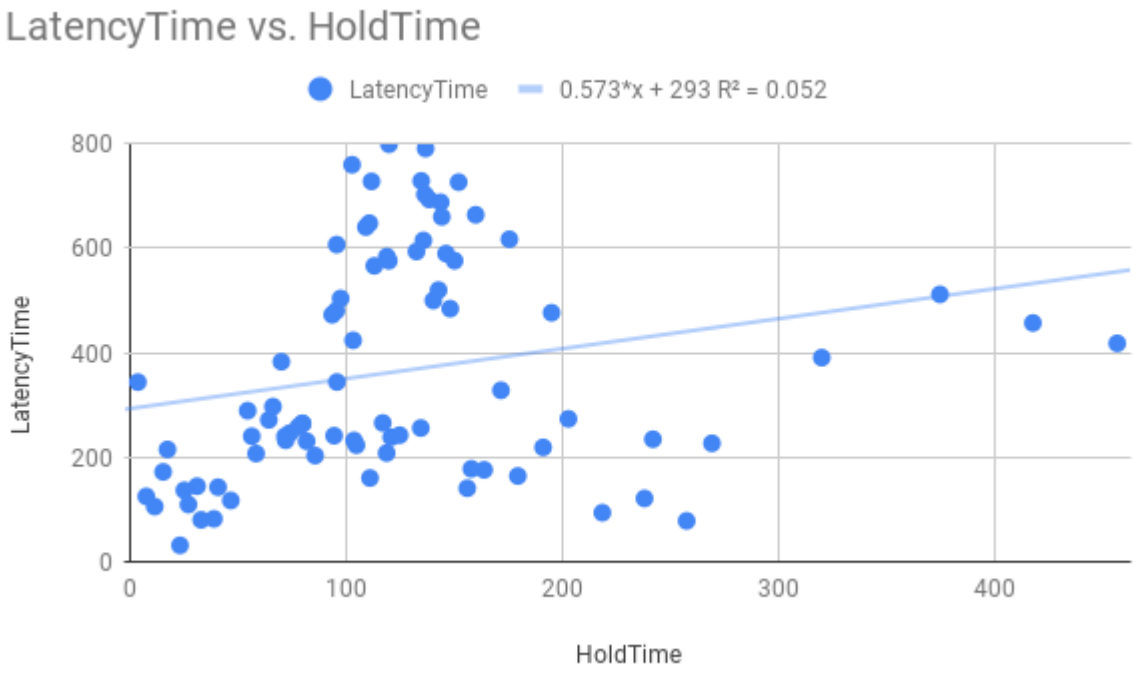


Figure 26

Correlation: 0.2271514258

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.573 times. LatencyTime and HoldTime are in positive correlation.

25. User Key – JHBOKKHOQW

BirthY-ear	Gender:	Parkins-ons:	Tremor-s	DiagnosisYear:	Sided:	UPDR-S:	Impact:	Levado-pa:	DA:	MAOB:	Other:
1939	Female	TRUE	True	2016	None	Don't know	Mild	TRUE	FALSE	FALSE	FALSE

Table 25

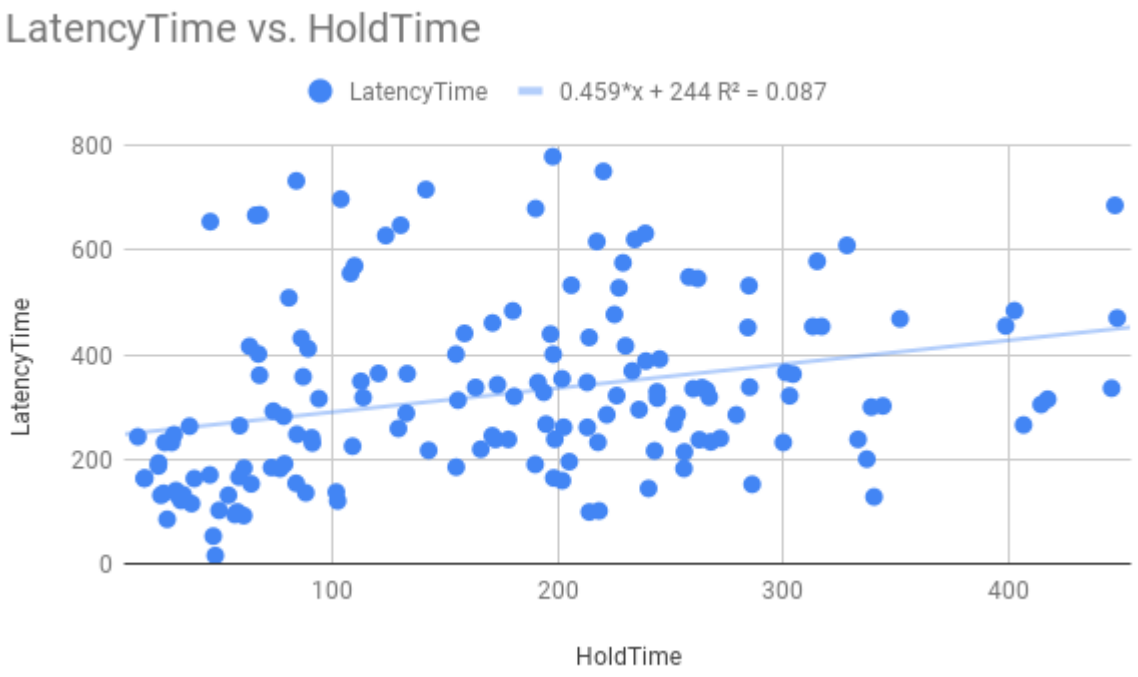


Figure 27

Correlation: 0.2941566124

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to increase by 0.459 times. LatencyTime and HoldTime are in positive correlation.

26. User Key – JMIVIRFQRR

BirthY-ear:	Gender:	Parkins-ons:	Tremor-s:	DiagnosisYear:	Sided:	UPDR-S:	Impact:	Levado-pa:	DA:	MAOB:	Other:
1947	Male	TRUE	TRUE	2007	Right	3	Severe	TRUE	TRUE	FALSE	FALSE

Table 26

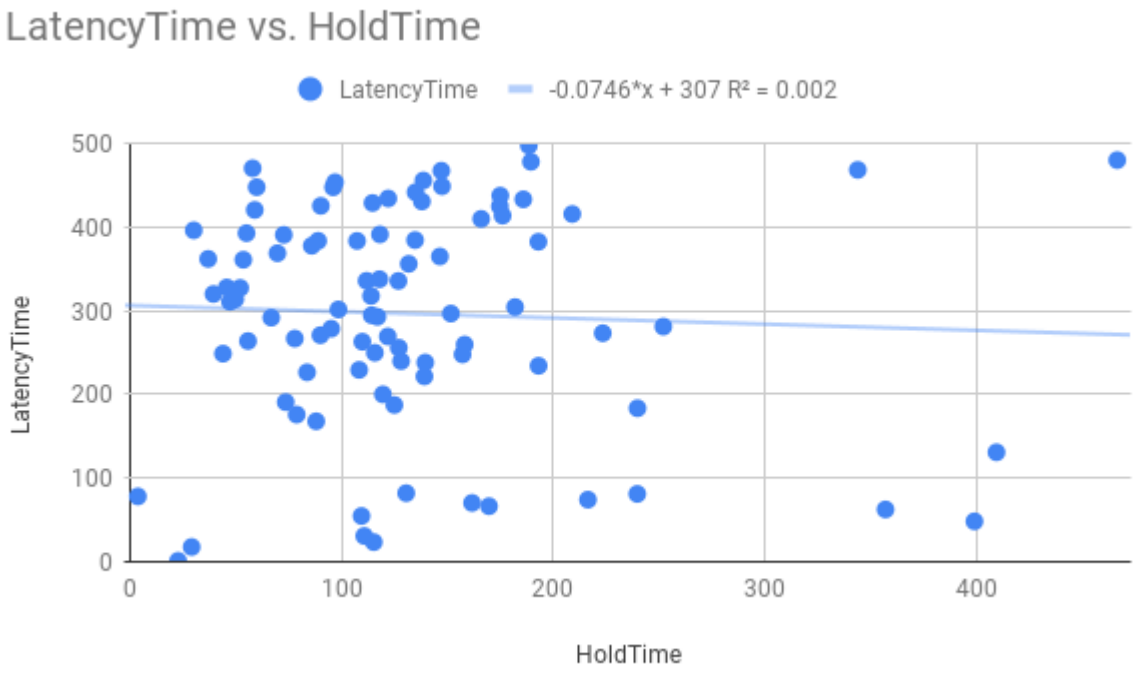


Figure 28

Correlation: -0.04736833829

For each decrease of one ms of HoldTime (Independent variable), the expected LatencyTime in ms is predicted to decrease by 0.0746 times. LatencyTime and HoldTime are in negative correlation.

27. User Key – LKR3NUMIFS

BirthY-ear:	Gender:	Parkins-ons:	Tremor-s:	DiagnosisYear:	Sided:	UPDR-S:	Impact:	Levado-pa:	DA:	MAOB:	Other:
--	Male	TRUE	TRUE	--	Right	Don't know	Mediu-m	TRUE	FALSE	FALSE	FALSE

Table 27

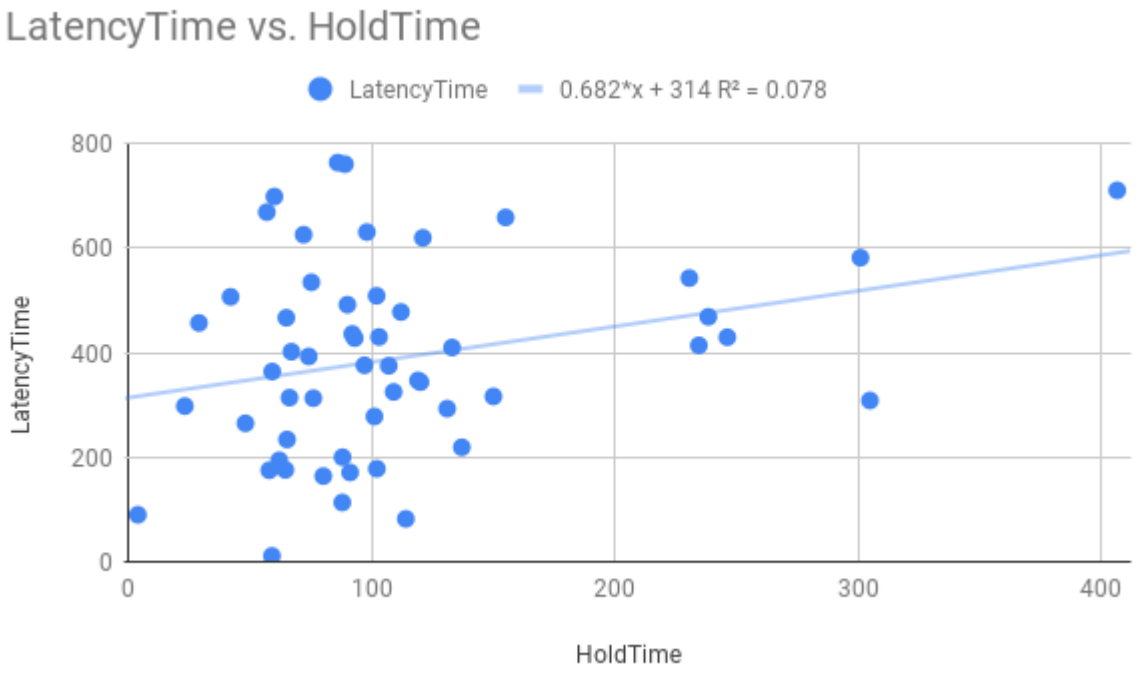


Figure 29

Correlation: 0.2799338816

For each increase of one ms of HoldTime (Independent variable), the expected LatencyTime (dependent) in ms is predicted to increase by 0.682 times. LatencyTime and HoldTime are in positive correlation.

3.4 Tools & Techniques used

We have used excel add-ins to find regression and correlation. CORREL function was used in excel and google sheets to find correlation. To find the regression curve we used a chart and scatter plot in google sheets. We also used python to find the regression and correlation curve. In python the libraries used for spearman and pearson we used scipy built-in functions pearson and spearman and for the linear regression we used a built-in library LinearRegression from sklearn.

CHAPTER - 4

Conclusion

From the analysis that we performed on the Tappy keystroke dataset from kaggle, we initially learnt to inspect the given data. We were able to find out details of the dataframe like the number of rows and columns, summary statistics of the given dataframe, data type of the columns in it, null values (if present), etc. We also learnt the way to find outliers in the data by plotting boxplot and we also became comfortable in plotting such boxplots. We found out how to extract any kind of data from the dataframe and how to store it in another file. We also became comfortable in dropping some unwanted data from the dataframe. We were also able to plot histograms for HoldTime and LatencyTime. We also became handy in merging some similar multiple files and performing similar analysis on them. Later we learnt to remove duplicates from specific columns which earlier used to create some problems related to correlation. Then while performing analysis on the data without duplicates, we became comfortable in taking out correlation in both python and excel. We also found some regression statistics for the particular file.

Lastly, we learnt to take out the linear regression graph in excel/google sheets as well in python from which we got to know about the type of correlation HoldTime and LatencyTime have and slope equation which we used in prediction of LatencyTime. Slope of the equation shows the intensity of the correlation between two variables.

Regarding the relationship we found between HoldTime and LatencyTime. Based on the impact of PD on a particular user the relationship between HoldTime & LatencyTime would vary. Further study can establish the actual relationship between the impacts of PD on the users along with intensity of relationship between both the variables.

References

- https://www.kaggle.com/valkling/tappy-keystroke-data-with-parkinsons-patients
- https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188226#sec008
- The dataset we worked on was taken from Kaggle.
- This particular dataset was made by Patrick DeKelly who is a Software Developer and Data Scientist with Deep Learning AI at Chelmsford, Massachusetts, United States.
- The research problem we chose was based on ‘High-accuracy detection of early Parkinson's
- Disease using multiple characteristics of finger movement while typing’ which was published by Warwick R. Adams - School of Computing & Mathematics, Charles Sturt University, N.S.W., Australia

Daily Activity Log

No.	Date	Duration	Task
01.	20-06-2021	3 hrs.	Watch and learn from the reference video.
02.	21-06-2021	3 hrs.	Got familiar with Python libraries and learned more about data science concepts.
03.	22-06-2021	4 hrs.	Learned about the dataset which we have to analyse and got to know about attributes which we have to analyse.
04.	23-06-2021	4 hrs.	We took a user input text file from tappy data of 2-3 kb size. We performed basic data inspection such as finding rows, columns, data types, summary of dataframe and no. of null values.
05.	24-06-2021	8 hrs.	We tried to create a group on Kaggle but it didn't work and learned what is outlier. We find out some outliers in hold time and latency time by plotting box plots.
06.	25-06-2021	5 hrs	We extracted outliers of hold time and latency time from the main file and stored them in another file and dropped outliers from the main file and saved it.
07.	26-06-2021	4 hrs.	From the archived user folder, we have selected such users which had valid birth year, tremors, levodopa, parkinsons, impact, gender etc. and found corresponding user data from tappy data for analysis.
08.	27-06-2021	5 hrs.	1. Picked up keystroke dataset for one userID, merged data from multiple keystroke txt files corresponding to that userID & labelled it as userID.csv file. 2. Took the above userID.csv file, performed the same data analysis as we did in the last 4 days. We took a snapshot/print of the code & resulted in a pdf/doc file.
09.	28-06-2021	8 hrs.	1. Plotted boxplot graph for that csv file & extracted outliers in separate CSV file. We kept normal data of the boxplot graph in a separate CSV file. 2. Saved everything in the shared google drive sub-folder.

			3. Repeated the process of 25th June and 26th June for at least 3 users.
10.	29-06-2021	8 hrs.	We tried to remove outliers by considering a threshold value of around 140ms to 150ms for holdtime and 600ms to 800ms for latency time. We were getting lot of duplicate records while doing this.
11.	30-06-2021	5 hrs.	So, we have been removing duplicates and then we will keep files which have greater than 50 records and less than 170 records for further actions and analysis.
12.	01-07-2021	5 hrs.	We found such files, which had decent records after removing duplicates. Then we started learning concepts of correlation and regression.
13.	02-07-2021	2 hrs.	We have learned about some extra functionalities and add-ons in Microsoft Excel.
14.	03-07-2021	3 hrs.	We performed Correlations using correl formula as well as add-ons method in Microsoft Excel. We also found out regression in excel.
15.	04-07-2021	4 hrs.	We tried to find out correlation and regression by using python.
16.	05-07-2021	4 hrs.	We tried taking out linear regression graphs for the user. We got to know the correct method to take out a linear regression plot and we got an intercept & slope formula.
17.	06-07-2021	5 hrs.	We plotted the correct linear regression graph for all our users and by using holdtime value in slope and intercept formula we predicted latency time. We found correlation by using another method that is by taking the square root of R square which we include beside the intercept & slope formula.
18.	07-07-2021	3 hrs.	We have merged all our output files in a single sheet with all plotted linear regression graphs and user info.
19.	08-07-2021	2 hrs.	Learning formal report writing.
20.	09-07-2021	2 hrs.	Collecting contents for report writing.
21.	10-07-2021	2 hrs.	Arranging sequence of topics in report.
22.	11-07-2021	2 hrs.	Formatting of report.
23.	12-07-2021	2 hrs.	Formatting of report.
24.	13-07-2021	2 hrs.	Adding additional components in the official report.
25.	14-07-2021	2 hrs.	Making the final official report ready with all components.

COMPLETION LETTER OF STUDENT

14th August, 2021

To, Faculty Coordinator

Subject: Completion letter of student.

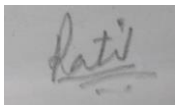
Dear Sir,

Upon a formal approval of you, Ashwati Shrimali from B.Tech Computer Science and Engineering has undergone Internship in DY Patil International University under my guidance.

She has successfully completed the summer internship for the project Neuroscience Data Analysis for a three to four weeks long internship programme.

She abided by the rules and regulations of the university and maintained a proper discipline with keen interest during the Internship.

Yours sincerely,



Ms. Priti Patil

STUDENT FEEDBACK OF INTERNSHIP (TO BE FILLED BY STUDENTS AFTER INTERNSHIP COMPLETION)

Student Name: Ashwati Shrimali Date: 14th August, 2021

Faculty Mentor: Ms. Priti Patil Title: Neuroscience-Data-Analysis

Mentor Email: priti.patil@dypiu.ac.in Internship is: Unpaid

Company/Organization: D.Y. Patil International University

Internship Address: Sector 29, Nigdi Pradhikaran, Akurdi, Pune, Maharashtra 411044

Faculty Coordinator: Dr. Kapil Sharma

Department: Computer Science

Dates of Internship: From 20/06/2021 To 14/08/2021

Please fill out the above in full detail

Give a brief description of your internship work (title and tasks for which you were responsible): Was your internship experience related to your major area of study?

Yes Yes, to a large degree Yes, to a slight degree No, not related at all.

Indicate the degree to which you agree or disagree with the following statements.

This experience has:	Strongly Agree	Agree	No Opinion	Disagree	Strongly Disagree
Given me the opportunity to explore a career field	Yes				
Allowed me to apply classroom theory to practice	Yes				
Helped me develop my decision-making and	Yes				

problem-solving skills					
Expanded my knowledge about the work world prior to permanent employment	Yes				
Helped me develop my written and oral communicatio n skills	Yes				
Provided a chance to use leadership skills (influence others, develop ideas with others, stimulate decision-making and action)	Yes				
This experience has:	Strongl-y Agree	Agree	No Opinio-n	Disagre e	Strongly Disagre e
Expanded my sensitivity to the ethical implication s of the work involved	Yes				

FACULTY MENTOR EVALUATION OF INTERN

Student Name: Ashwati Shrimali Date:14th August 2021

Faculty Mentor: Ms. Priti Patil Title: Neuroscience-Data-Analysis

Company/Organization: D.Y. PATIL INTERNATIONAL UNIVERSITY

Internship Address: Sector 29, Nigdi Pradhikaran, Akurdi, Pune, Maharashtra 411044

Dates of Internship: From 20/06/2021 To 14/07/2021

Please evaluate your intern by indicating the frequency with which you observed the following behaviors:

Parameters	Needs improvement	Satisfactory	Good	Excellent
Behaviors			✓	
Performs in a dependable manner				✓
Cooperates with co-workers and supervisors				✓
Shows interest in work			✓	
Learns quickly			✓	
Shows initiative				✓
Produces high quality work			✓	
Accepts responsibility			✓	
Accepts criticism		✓		
Demonstrates organizational skills		✓		
Uses technical knowledge and expertise				✓
Shows good judgment			✓	
Demonstrates creativity/originality		✓		
Analyzes problems effectively				✓
Is self-reliant				✓
Communicates well			✓	
Writes effectively		✓		
Has a professional attitude			✓	
Gives a professional appearance			✓	

Is punctual				✓
Uses time effectively				✓

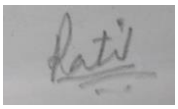
Overall performance of student intern (highlight one):

(Needs improvement/ Satisfactory/ Good/ Excellent)

Additional comments, if any:

She could follow the instructions properly & could apply those appropriately.

Signature of Mentor



(Ms. Priti Patil)