A PROJECT REPORT ON

PROJECT 1:

EXPERENTIAL DATA COLLECTION

AND

DESCRIPTIVE STATISTICAL ANALYSIS

BY

UTA MAV ID	NAME
1002050821	ABHISHEK SHRINIVAS JOSHI

UNDER THE GUIDANCE OF

PROF DR. EMMA YANG

OCTOBER 20, 2022

I Abhishek Shrinivas Joshi did not give or receive any assistance on this project, and the report submitted is wholly my own.

CONTENTS

INTRODUCTION

The project's goals are to collect data and run a chi-square goodness of fit test on the information that was acquired for project I. For the study, two datasets were obtained: one physically from the UTA Sustainable and Intelligent Manufacturing (SIGMA) Lab, and the other virtually from the American Airlines website after being enhanced to the required format.

This project report details the data collected as well as the outcomes of the associated descriptive statistical analysis. Finding out if Data Set I follows a normal distribution and Data Set II follows an exponential distribution is the major purpose of the study.

The maximum diameter of Nylon 12 powder particles selected at random for 3D printing is observed in 125 times in Data Set I. In this case, I employ the maximum diameter, a continuous random variable that is assumed to have a normal distribution.

Data Set II contains 104 observations of the time differences between flights arriving at JFK airport on November 29, 2022, between one and nine in the morning.

DATA COLLECTION

Data Set I:

On October 7, 2022, at 11 a.m., a time slot was designated for the data collection of the maximum diameter of randomly selected Nylon 12 particles. Our GTA Mr. Ahnaf Shahriar directed us as we carried out this activity in Engineering Research Building 188.

Procedure:

1. Set a sample of fresh, used, or recycled powder on the microscope's stage.
2. Move the microscope to an area where the particles are less tightly packed.
3. To prevent the particles on the screen from appearing to be vibrating while you take a screenshot of them, click the Vibration Disable option.
4. Using the screenshot as a reference, we performed several dark mode measurements. The particles in the screenshot were the ones with red colours.
5. Next, the CSV file containing the parameters that were measured was exported and stored.
6. For the sake of statistical analysis, I will only consider the maximum diameter as a random variable.

Notes:

1. The particles with noticeably smaller sizes would be considered to have a maximum diameter of zero.
2. Since two particles behave as one when they are too close to one another, we can also think of those particles as outliers.

Data Set II:

For collecting the second dataset we used **American Lines official website**.

Procedure:

1. Dataset contains three columns
    - Flight Number: Fight number of flight which landed
    - Arrival Time: Time stamp of the arrival time of the flight
    - Interval Time (min): time difference between flights arriving
2. The random variable X, which is the difference between the timestamps of two subsequent flight arrivals at JFK airport on November 29, 2022.

DESCRIPTIVE STATISTICAL ANALYSIS

Dataset I:

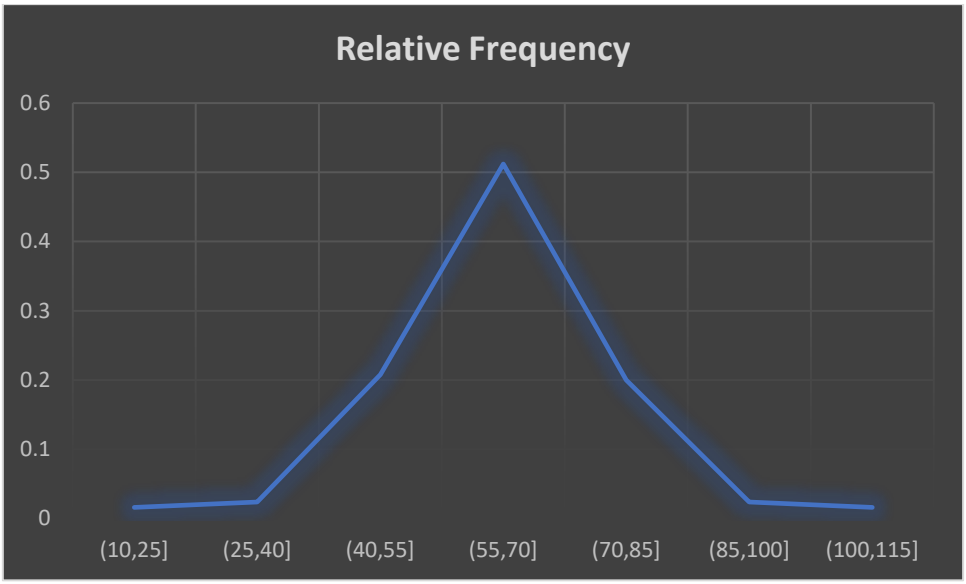
The total number of observations for Dataset I (n): 125

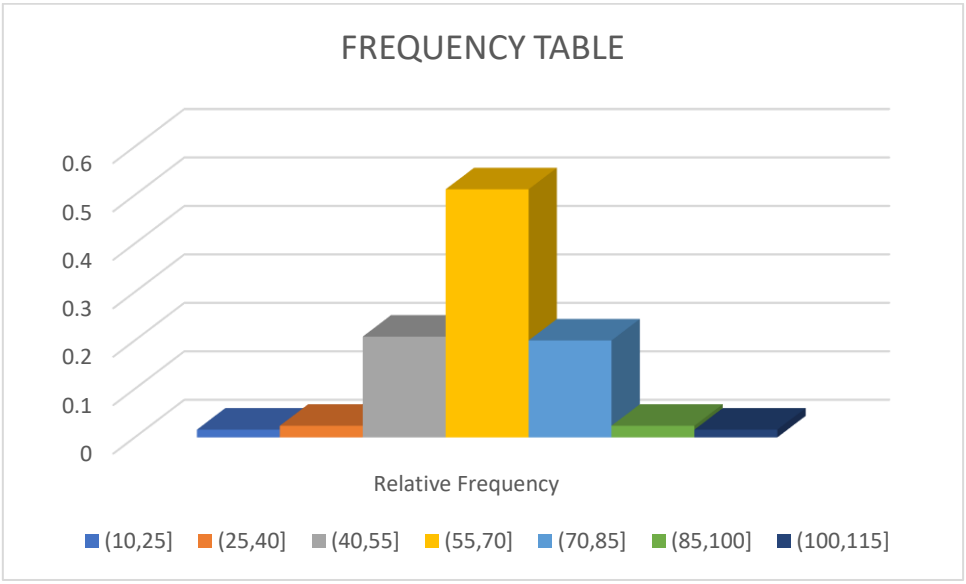
Measures of Location	
Mean - μm	63.72672
Median- μm	64.45
Mode- μm	60.18

Measures of Variability	
Range- μm	85.64(110.27-24.63)
Variance- μm	187.4109196
Standard Deviation- μm	13.7449008
Coefficient of Variation- μm	0.21568505

Percentiles	
Q1	55.48
Q2	64.45
Q3	69.49
IQR	14.01

Tabular Summary			
Class Interval	**Count or Frequency**	**Relative Frequency**	**Cumulative Relative Frequency**
(10,25]	2	0.016	0.016
(25,40]	3	0.024	0.04
(40,55]	26	0.208	0.248
(55,70]	64	0.512	0.76
(70,85]	25	0.2	0.96
(85,100]	3	0.024	0.984
(100,115]	2	0.016	1


Relative Frequency
0.6
0.5
0.4
0.3
0.2
0.1
0
(10,25]
(25,40]
(40,55]
(55,70]
(70,85]
(85,100]
(100,115]


FREQUENCY TABLE
0.6
0.5
0.4
0.3
0.2
0.1
0
Relative Frequency
(10,25]
(25,40]
(40,55]
(55,70]
(70,85]
(85,100]
(100,115]

Dataset II:

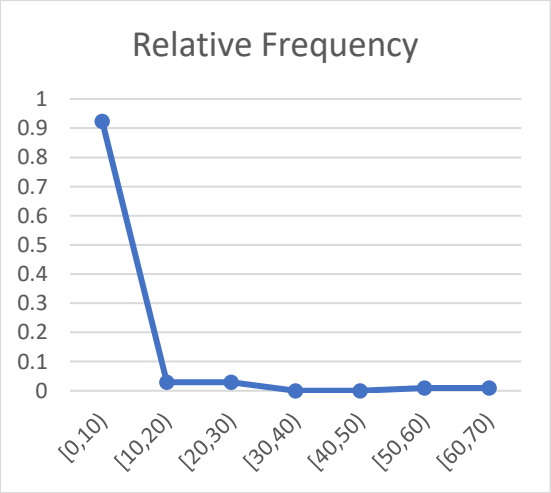
The total number of observations for Dataset II (n): 104

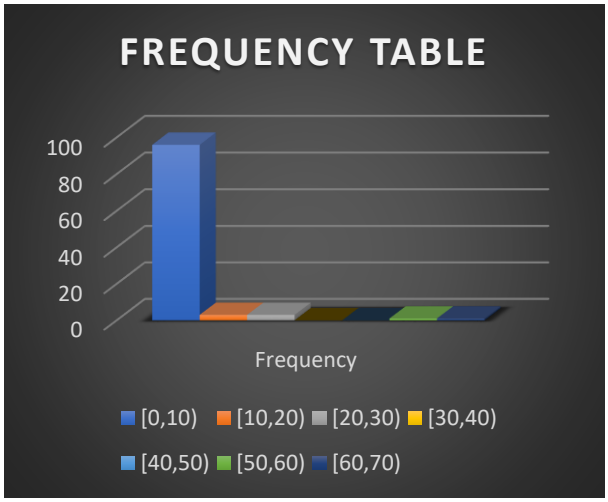
Measures of Location	
Mean – µm	4.548076923
Median- µm	2
Mode- µm	2

Measures of Variability	
Range- µm	60 (60-0)
Variance- µm	70.9399963
Standard Deviation- µm	8.463376048
Coefficient of Variation- µm	1.860869152

Percentiles	
Q1	2
Q2	2
Q3	4
IQR	2

Tabular Summary			
Class Interval	Count or Frequency	Relative Frequency	Cumulative Relative Frequency
[0,10)	96	0.923076923	0.923076923
[10,20)	3	0.028846154	0.951923077
[20,30)	3	0.028846154	0.980769231
[30,40)	0	0	0.980769231
[40,50)	0	0	0.980769231
[50,60)	1	0.009615385	0.990384615
[60,70)	1	0.009615385	1





CHI-SQUARE GOODNESS-OF-FIT-TESTS

I utilized the data to do the Chi-square analysis for Data Sets I and II. To do this, I went through the steps below.

Procedure:

1. I divided the data into various classes.
2. Next, I used the built-in NORM.DIST () excel function for Data Set I and the GAMMA.DIST () excel function for Data Set II to calculate the Class Frequency (Oi) for each of the partitions.
3. After that, I multiplied the relevant class probabilities by the total number of observations to arrive at the expected class frequencies.
4. Next, I changed the class sizes to make sure that each class was expected to meet more frequently than four times (ei).
5. Next, for all classes, I discovered the observed frequency (oi). The chi-square class component for each class was then determined from there.
6. The total of the observed, class, expected frequencies, and chi-square class components was then discovered.
7. Next, using the decision rule "Reject Ho when calculated chi-square is greater than tabular one," I calculated the chi-squared test statistic, which is the sum of all chi-squared class components, and compared it with the chi-squared table value with a significance level of o.o5 and k-1 degrees of freedom.
8. Keep in mind that accepting Ho is a tentative conclusion and that it does not offer statistically significant proof that the facts fit the theoretical distribution.

Data Set I:

Class	**Observed Frequency (o_i)**	**Class Probability**	**Class Expected Frequency (e_i)**	**Chi square Class Component**
X<=40	5	0.042153938	5.269242255	0.01375746
40<X<=55	26	0.220590923	27.57386544	0.089833339
55<X<=70	64	0.413206761	51.65084515	2.9525485
70<X<=85	25	0.263203709	32.9004636	1.897156401
X>85	5	0.060844668	7.605583552	0.892642307
Total	125	1	125	5.845938008
From Table	α=0.05	v=n-1=5-1=4		9.488

Chi-square Test Statistic	5.845938008
Chi-square(α=0.05, v=4)	9.488

The following setup will be used to determine whether Data Set I has a normal distribution:

H_o: Data Set I follows normal distribution.

$H_{1:}$ Data Set I does not follow normal distribution.

Conclusion: At 5% significant level,

We found that Chi-square Test Statistic < Chi-square (α=0.05, v=4),

Hence we failed to reject the null hypothesis.

Interpretation: Thus Data Set I follows normal distribution.

Data Set II:

Class	**Observed Frequency (oi)**	**Class Probability**	**Class Expected Frequency (ei)**	**Chi square Class Component**
X<4	72	0.585006574	60.84068372	2.046826765
4<=X<8	21	0.242773882	25.24848376	0.714879137
8<X<=12	5	0.100749565	10.47795477	2.863916588
X>12	6	0.071469978	7.432877744	0.27622392
Total	104	1	104	5.90184641
From Table	α=0.05	v=n-1=4-1=3		7.815

Chi-square Test Statistic	5.90184641
Chi-square(α=0.05, v=4)	7.815

The following setup will be used to determine whether Data Set II has an exponential distribution:

H_o: Data Set II follows exponential distribution.

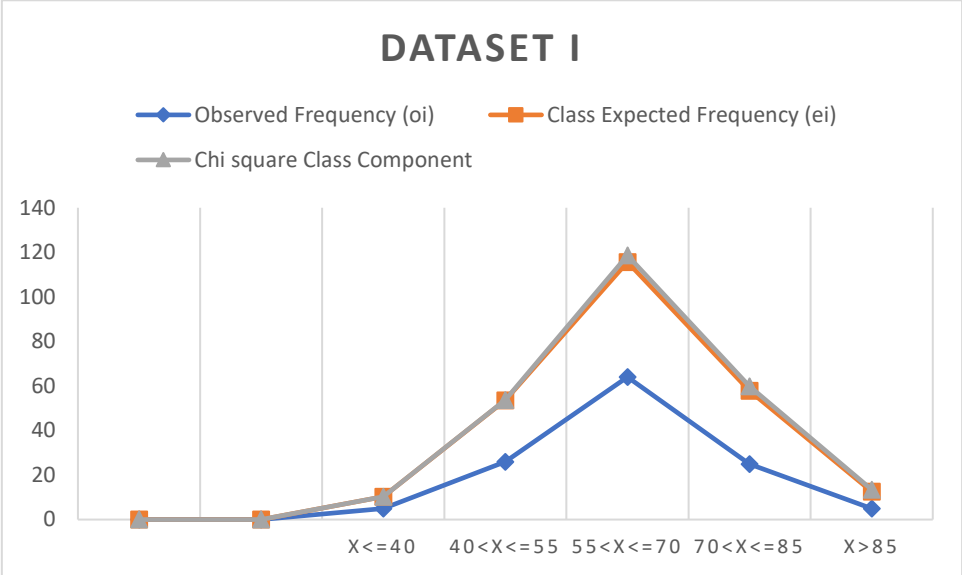
$H_{1:}$ Data Set II does not follow exponential distribution.

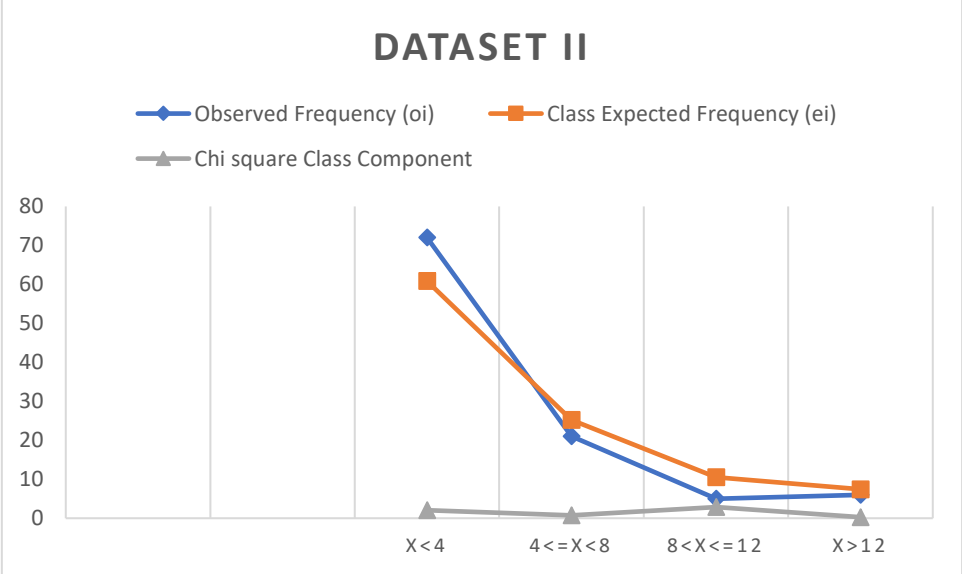
Conclusion: At 5% significant level,

We found that Chi-square Test Statistic < Chi-square (α=0.05, v=3),

Hence we failed to reject the null hypothesis.

Interpretation: Thus Data Set II follows exponential distribution.


DATASET I
Observed Frequency (oi)
Class Expected Frequency (ei)
Chi square Class Component
140
120
100
80
60
40
20
0
X<=40
40<X<=55
55<X<=70
70<X<=85
X>85


DATASET II
Observed Frequency (oi)
Class Expected Frequency (ei)
Chi square Class Component
80
70
60
50
40
30
20
10
0
X<4
4<=X<8
8<X<=12
X>12

CONCLUSION

I draw the conclusion that Dataset I may have a normal distribution from these tests because we were unable to rule out the null hypothesis. Because adopting the null hypothesis is a tenuous conclusion, we cannot state with a high degree of assurance that dataset one follows a normal distribution.

Also I draw same conclusion that Dataset II may have an exponential distribution from these tests because we were unable to rule out the null hypothesis. Because adopting the null hypothesis is a tenuous conclusion, we cannot state with a high degree of assurance that dataset two follows an exponential distribution.

APPENDIX

DATASET I:

Class	Observed Frequency (o_i)	Class Probability	Class Expected Frequency (e_i)	Chi square Class Component
X<=40	5	0.042153938	5.269242255	0.01375746
40<X<=55	26	0.220590923	27.57386544	0.089833339
55<X<=70	64	0.413206761	51.65084515	2.9525485
70<X<=85	25	0.263203709	32.9004636	1.897156401
X>85	5	0.060844668	7.605583552	0.892642307
Total	125	1	125	5.845938008
From Table	α=0.05	v=n-1=5-1=4		9.488

Class	Class Probability
X<=40	NORM.DIST(40, 63.72672, 13.7449008,TRUE)
40<X<=55	NORM.DIST(55, 63.72672, 13.7449008,TRUE)-NORM.DIST(40, 63.72672, 13.7449008,TRUE)
55<X<=70	NORM.DIST(70, 63.72672, 13.7449008,TRUE)-NORM.DIST(55, 63.72672, 13.7449008,TRUE)
70<X<=85	NORM.DIST(85, 63.72672, 13.7449008,TRUE)-NORM.DIST(70, 63.72672, 13.7449008,TRUE)
X>85	1-NORM.DIST(85, 63.72672, 13.7449008,TRUE)

DATASET II:

Class	Observed Frequency (oi)	Class Probability	Class Expected Frequency (ei)	Chi square Class Component
X<4	72	0.585006574	60.84068372	2.046826765
4<=X<8	21	0.242773882	25.24848376	0.714879137
8<X<=12	5	0.100749565	10.47795477	2.863916588
X>12	6	0.071469978	7.432877744	0.27622392
Total	104	1	104	5.90184641
From Table	α=0.05	v=n-1=4-1=3		7.815

Class	Class Probability
X<4	GAMMA.DIST(4,1,4.55,TRUE)
4<=X<8	GAMMA.DIST(8,1,4.55,TRUE)-GAMMA.DIST(4,1,4.55,TRUE)
8<X<=12	GAMMA.DIST(12,1,4.55,TRUE)-GAMMA.DIST(8,1,4.55,TRUE)
X>12	1-GAMMA.DIST(12,1,4.55,TRUE)

CHI-SQUARE CLASS COMPONENT : $\frac{(Oi-ei)\text{^}2}{ei}$

DATASET 1 OR DATASET I:

SR.NO	Population
	Maximum Diameter
1	49.11
2	75.64
3	64.23
4	53.75
5	63.92
6	53.94
7	67.61
8	75.52
9	79.07
10	68.67
11	68.15
12	59.6
13	58.09
14	68.45
15	68.3
16	48.63
17	70.69
18	64.45
19	65.67
20	82.94
21	68.81
22	40.29
23	80.2
24	67.94
25	58.85
26	64
27	59.61
28	54.28
29	75.25
30	67.14
31	68.93
32	79.2

33	68.54
34	50.78
35	58.12
36	68.05
37	65.02
38	66.45
39	70
40	60.09
41	81.43
42	47.28
43	28.53
44	55.48
45	67.15
46	55
47	57.83
48	60.61
49	56.12
50	59.66
51	71.22
52	67.48
53	59.36
54	51.87
55	61.04
56	51.04
57	54.71
58	24.7
59	84.71
60	66.46
61	51.52
62	50.38
63	68.6
64	60.18
65	59.27
66	75.14
67	81.4
68	43.88
69	24.63
70	46.36
71	83.46
72	66.05
73	51.29
74	72.44
75	60.18
76	60.78
77	38.78
78	48.98

79	63.69
80	57.14
81	77.34
82	80.85
83	54.48
84	61.42
85	56.54
86	66.9
87	65.35
88	98.11
89	63.41
90	79.28
91	48.83
92	72.55
93	82.7
94	50.54
95	62.53
96	57.94
97	73.23
98	65.65
99	66.97
100	54.08
101	95.33
102	42.78
103	80.42
104	59.56
105	49.65
106	66.81
107	65.35
108	70.85
109	54.69
110	47.37
111	103.6
112	64.75
113	71.82
114	61.36
115	64.81
116	110.27
117	60.74
118	69.49
119	80.34
120	56.79
121	37.44
122	85.02
123	67.3
124	68.57

125	68.22

DATASET 2 OR DATASET II:

FlightNo.	Arrival Time	Interval Time(in min)
A320 (N516JB)	1:04	0
B748 (LX-VCE)	1:56	52
B763 (N617AZ)	2:00	4
A320 (N686TA)	2:08	8
B763 (N359AZ)	2:11	3
B738 (HP-1838CMP)	2:32	21
A321 (N903JB)	3:32	60
B190 (C-GSKG)	3:36	4
B38M (9Y-TTO)	4:05	29
A321 (N967JT)	4:18	13
A306 (N163UP)	4:26	8
A320 (N809JB)	4:29	3
MD11 (N273UP)	4:36	7
A21N (N2043J)	4:38	2
A320 (N992AV)	4:40	2
B763 (N178DN)	4:42	2
A321 (N942JB)	4:43	1
A333 (N801NW)	4:50	7
B763 (G-DHLL)	4:51	1
B763 (N184DN)	4:53	2
A21N (N2044J)	4:55	2
A320 (N703JB)	5:02	7

A339 (N419DX)	5:06	4
A21N (N2086J)	5:13	7
A20N (N934AG)	5:16	3
B789 (4X-EDE)	5:19	3
A320 (N659JB)	5:28	9
A320 (N807JB)	5:31	3
A320 (N607JB)	5:33	2
B763 (N329UP)	5:36	3
B772 (N776AN)	5:36	0
A321 (N971JT)	5:38	2
B752 (N722TW)	5:38	0
B763 (N177DN)	5:42	4
A306 (N748FD)	5:45	3
A321 (N979JT)	5:46	1
B77W (N721AN)	5:48	2
A321 (N109NN)	5:50	2
B763 (N374UP)	5:51	1
B739 (N915DU)	5:52	1
A359 (B-LRQ)	5:56	4
B38M (N306RC)	5:58	2
A321 (N972JT)	6:00	2
B739 (N881DN)	6:02	2
B38M (N338RS)	6:05	3
A320 (N597JB)	6:07	2
CRJ9 (N228PQ)	6:09	2
CRJ9 (N176PQ)	6:11	2
B739 (N285AK)	6:12	1
E75S (N117HQ)	6:14	2
A321 (N946JL)	6:15	1
CRJ9 (N153PQ)	6:17	2

B39M (N920AK)	6:18	1
A321 (N980JT)	6:20	2
CRJ9 (N937XJ)	6:22	2
E190 (N187JB)	6:23	1
B38M (N335RT)	6:28	5
A321 (N111ZM)	6:32	4
B739 (N814DN)	6:34	2
CRJ9 (N691CA)	6:38	4
A321 (N977JE)	6:39	1
B752 (N947FD)	6:43	4
B752 (N690DL)	6:44	1
B77W (VT-ALN)	6:46	2
MD11 (N573FE)	6:48	4
E75S (N246JQ)	6:51	3
A333 (D-AJFK)	6:52	1
B763 (N1605)	6:56	4
E190 (N348JB)	6:56	0
E190 (N373JB)	6:58	2
A320 (N503JB)	7:00	2
E75S (N108HQ)	7:02	2
E75S (N242JQ)	7:04	4
BCS1 (N103DU)	7:04	0
E75S (N113HQ)	7:05	1
A320 (N768JB)	7:07	2
CRJ9 (N326PQ)	7:09	2
CRJ9 (N301PQ)	7:11	2
B763 (N362CM)	7:12	1
E190 (N358JB)	7:14	2
B789 (CC-BGL)	7:15	1
CRJ9 (N582CA)	7:19	4

A332 (N382HA)	7:20	1
B39M (N949AK)	7:23	3
A321 (N996JL)	7:25	2
E75S (N233JQ)	7:27	2
B739 (N275AK)	7:28	1
B77W (N729AN)	7:30	2
A321 (N106NN)	7:32	2
B772 (N765AN)	7:34	2
E190 (N203JB)	7:36	2
A321 (N973JT)	7:37	1
A321 (N957JB)	7:39	2
CRJ9 (N320PQ)	7:41	2
B752 (N723TW)	7:43	2
A320 (N657JB)	7:43	0
A320 (N537JT)	7:47	4
B772 (N790AN)	7:54	7
E190 (N307JB)	7:56	2
B764 (N832MH)	8:06	10
B763 (CC-CWV)	8:16	10
A388 (A6-EUR)	8:23	7
A21N (N924VA)	8:30	7
A320 (N828JB)	8:53	23

REFERENCE

1. DASC 5302 – PROJECT DESCRIPTION – 1 and 2.
2. TA RECORDING
3. https://docs.google.com/spreadsheets/d/1iwbq1Oed5lixsyWM-TZaevz3Ygn3sM4h/edit?usp=sharing&ouid=101786930800661555462&rtpof=true&sd=true
4. Dataset II : https://www.flightradar24.com/data/airports/jfk/arrivals