



Diplomado en Ciencia de Datos UNAM

Modulo 13 Uso de Spark Y Hadoop

Alumno: Sergio Ibarra Ramírez



Contenido

1. *Descargar varios archivos referente a un tema en particular*
2. *Cargar los archivos al cluster de HADOOP*
3. *Crear un Data Frame en Spark*
4. *Hacer cambios al data frame (columnas o renglones)*
5. *Guarda el nuevo DF en HDFS y en el Bucket*
6. *Agregar evidencia de que el JOB se ejecuto en el cluster*

¿Qué hay disponible al momento en mi cluster?

```
Last login: Fri Sep 22 04:30:34 2023 from 35.235.244.32
sergio_ibarra1795@cluster-sir-dcd05-m:~$ ls -la
total 142408
drwxr-xr-x 3 sergio_ibarra1795 sergio_ibarra1795 4096 Sep  9 19:46 .
drwxr-xr-x 4 root               root          4096 Sep  9 17:45 ..
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 1147 Sep 22 02:17 .bash_history
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 220 Mar 27 2022 .bash_logout
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 3526 Mar 27 2022 .bashrc
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 807 Mar 27 2022 .profile
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 1409 Sep  9 19:46 .python_history
drwx-r--r-- 2 sergio_ibarra1795 sergio_ibarra1795 4096 Sep 22 04:32 .ssh
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 2226045 Jan 17 2021 2000-0.txt
-rw-r--r-- 1 sergio_ibarra1795 sergio_ibarra1795 143562785 Sep  9 18:35 linea-mujeres-cdmx.csv
sergio_ibarra1795@cluster-sir-dcd05-m:~$
```

```
sergio_ibarra1795@cluster-sir-dcd05-m:~$ hdfs dfs -ls -R /tmp
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 19:52 /tmp/dcd
drwxr-xr-x - root hadoop 0 2023-09-09 19:52 /tmp/dcd/ecobici
-rw-r--r-- 2 root hadoop 41926344 2023-09-09 19:52 /tmp/dcd/ecobici/2023-01.csv
-rw-r--r-- 2 root hadoop 47532884 2023-09-09 19:52 /tmp/dcd/ecobici/2023-02.csv
-rw-r--r-- 2 root hadoop 58449187 2023-09-09 19:52 /tmp/dcd/ecobici/2023-03.csv
-rw-r--r-- 2 root hadoop 58119473 2023-09-09 19:52 /tmp/dcd/ecobici/2023-04.csv
-rw-r--r-- 2 root hadoop 67532130 2023-09-09 19:52 /tmp/dcd/ecobici/2023-05.csv
-rw-r--r-- 2 root hadoop 72151312 2023-09-09 19:52 /tmp/dcd/ecobici/2023_06.csv
-rw-r--r-- 2 root hadoop 71416339 2023-09-09 19:52 /tmp/dcd/ecobici/2023_07.csv
-rw-r--r-- 2 root hadoop 79235355 2023-09-09 19:52 /tmp/dcd/ecobici/2023_08.csv
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 18:39 /tmp/dcd/sirilo
-rw-r--r-- 2 sergio_ibarra1795 hadoop 143562785 2023-09-09 18:39 /tmp/dcd/sirilo/linea-mujeres-cdmx.csv
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 19:33 /tmp/dcd/wordcount
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 18:03 /tmp/dcd/wordcount/input
-rw-r--r-- 2 sergio_ibarra1795 hadoop 2226045 2023-09-09 18:03 /tmp/dcd/wordcount/input/texto.txt
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 18:09 /tmp/dcd/wordcount/output
-rw-r--r-- 2 sergio_ibarra1795 hadoop 0 2023-09-09 18:09 /tmp/dcd/wordcount/output/_SUCCESS
-rw-r--r-- 2 sergio_ibarra1795 hadoop 148653 2023-09-09 18:09 /tmp/dcd/wordcount/output/part-r-00000
-rw-r--r-- 2 sergio_ibarra1795 hadoop 149555 2023-09-09 18:09 /tmp/dcd/wordcount/output/part-r-00001
-rw-r--r-- 2 sergio_ibarra1795 hadoop 148719 2023-09-09 18:09 /tmp/dcd/wordcount/output/part-r-00002
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 19:33 /tmp/dcd/wordcount/pyspark
-rw-r--r-- 2 sergio_ibarra1795 hadoop 0 2023-09-09 19:33 /tmp/dcd/wordcount/pyspark/_SUCCESS
-rw-r--r-- 2 sergio_ibarra1795 hadoop 324685 2023-09-09 19:33 /tmp/dcd/wordcount/pyspark/part-00000
-rw-r--r-- 2 sergio_ibarra1795 hadoop 321210 2023-09-09 19:33 /tmp/dcd/wordcount/pyspark/part-00001
drwxrwxrwt - hdfs hadoop 0 2023-09-09 16:36 /tmp/hadoop-yarn
drwxrwxrwt - hdfs hadoop 0 2023-09-09 18:08 /tmp/hadoop-yarn/staging
drwxrwxrwt - hdfs hadoop 0 2023-09-09 16:36 /tmp/hadoop-yarn/staging/history
drwx----- - sergio_ibarra1795 hadoop 0 2023-09-09 18:08 /tmp/hadoop-yarn/staging/sergio_ibarra1795
drwx----- - sergio_ibarra1795 hadoop 0 2023-09-09 18:09 /tmp/hadoop-yarn/staging/sergio_ibarra1795/
.staging
drwx-wx-wx - hive hadoop 0 2023-09-09 16:37 /tmp/hive
drwx-wx-wx - hive hadoop 0 2023-09-22 04:28 /tmp/hive/_resultscache_
drwx----- - hive hadoop 0 2023-09-22 04:28 /tmp/hive/_resultscache_/results-b8f62619-0
043-4a0b-a700-bd67bf9dba42
```



Descargar varios archivos referente a un tema en particular



The screenshot shows a Jupyter Notebook interface with the title "Mod13_Uso_Pyspark_tarea_sesion1 (unsaved changes)". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for saving, adding, deleting, and running cells, as well as a "Run" button and a "Markdown" dropdown. The notebook content consists of two text cells and one code cell. The first text cell says "Intentamos bajar la data de los archivos de British election study". The second text cell says "Intentamos bajar la data de los archivos de British election study Part3". The code cell contains a terminal output of a `wget` command downloading a CSV file from data.world. The output shows the file being resolved, connected to, and saved successfully.

```
jupyter Mod13_Uso_Pyspark_tarea_sesion1 (unsaved changes)
```

File Edit View Insert Cell Kernel Widgets Help Trusted | PySpark

Intentamos bajar la data de los archivos de British election study

Intentamos bajar la data de los archivos de British election study Part3

```
In [1]: ! wget https://data.world/wwymak/british-election-study-part-3-csv-files/workspace/file?filename=BES2017_W13_Panel_v1.0-3.csv
```

```
--2023-09-22 06:47:36-- https://data.world/wwymak/british-election-study-part-3-csv-files/workspace/file?filename=BES2017_W13_Panel_v1.0-3.csv
Resolving data.world (data.world)... 151.101.66.133, 151.101.2.133, 151.101.130.133, ...
Connecting to data.world (data.world)|151.101.66.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 32598 (32K) [text/html]
Saving to: 'file?filename=BES2017_W13_Panel_v1.0-3.csv.1'

file?filename=BES20 100%[=====>] 31.83K --.-KB/s in 0.003s

2023-09-22 06:47:37 (9.09 MB/s) - 'file?filename=BES2017_W13_Panel_v1.0-3.csv.1' saved [32598/32598]
```

Borramos data que habíamos cargado anteriormente para solo tener archivos de British election study

Cargar los archivos al cluster de HADOOP

In []: Se tuvo que renombrar los archivos de csv de British election para lograr un mejor manejo

In [10]: `import os`

```
# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
file_path = os.path.join(cwd, "file?filename=BES2017_W13_Panel_v1.0-3.csv.1")

# Rename the file.
os.rename(file_path, "bes2017_part3.csv")
```

In []: Y se intenta subir de nuevo el archivo a la carpeta british dentro de mi cluster de Hadoop

In [11]: `! hdfs dfs -put bes2017_part3.csv /tmp/dcd/british/`

In []: Se verifica que el archivo de nombre bes2017_part3.csv se haya subido correctamente a /tmp/dcd/british/

In [12]: `! hdfs dfs -ls -R /tmp/dcd/british`

```
-rw-r--r--  2 root hadoop    32598 2023-09-22 07:09 /tmp/dcd/british/bes2017_part3.csv
```

```
# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
file_path = os.path.join(cwd, "british-election-study-part-2-csv-files")

# Rename the file.
os.rename(file_path, "bes2017_part2.csv")
```

In []: Y se intenta subir de nuevo el archivo bes2017_part1.csv a la carpeta british dentro de mi cluster de Hadoop

In [19]: `! hdfs dfs -put bes2017_part2.csv /tmp/dcd/british/`

In [20]: `! hdfs dfs -ls -R /tmp/dcd/british`

```
-rw-r--r--  2 root hadoop    39212 2023-09-22 07:14 /tmp/dcd/british/bes2017_part1.csv
-rw-r--r--  2 root hadoop    39606 2023-09-22 07:15 /tmp/dcd/british/bes2017_part2.csv
-rw-r--r--  2 root hadoop    32598 2023-09-22 07:09 /tmp/dcd/british/bes2017_part3.csv
-rw-r--r--  2 root hadoop    32598 2023-09-22 07:13 /tmp/dcd/british/bes2017_part4.csv
```

Crear un Data Frame en Spark

3. Crear un Data Frame en Spark

```
In [32]: british2_df = spark.read.csv('hdfs:///tmp/dcd/british2/', header=True, inferSchema=True)
```

```
In [33]: british2_df.show(10)
```

```
-----+
|<!doctype html><html lang="en"><head> <title data-react-helmet="true">data.world</title>    <meta charset="utf-8"><meta name
="viewport" content="width=device-width|initial-scale=1"><meta name="referrer" content="origin-when-cross-origin"><meta name
="og:image" content="https://d2ogkq1rg66kh0.cloudfront.net/site-resources/images/fb-image-share.3a87a6eb.png"><meta name="og:
site_name" content="data.world"><meta name="twitter:card" content="summary"><meta name="twitter:site" content="@datadotworl
d"><meta property="fb:app_id" content="776416075825233"><link rel="stylesheet" href="https://fonts.googleapis.com/css2?family
=Inconsolata:wght@400;700&family=Lato:ital|wght@0|300;0|400;0|700;1|300;1|400;1|700&display=swap"><link rel="search" href="/o
pensearch.xml" type="application/opensearchdescription+xml" title="data.world">  <link rel="apple-touch-icon" sizes="57x57" h
ref="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-57x57.png"><link rel="apple-touch-ico
n" sizes="60x60" href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-60x60.png"><link rel
="apple-touch-icon" sizes="72x72" href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-72x
72.png"><link rel="apple-touch-icon" sizes="76x76" href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/app
le-touch-icon-76x76.png"><link rel="apple-touch-icon" sizes="114x114" href="https://assets.data.world/icons-d1940ad25da01d05b
8f27389101617c1/apple-touch-icon-114x114.png"><link rel="apple-touch-icon" sizes="120x120" href="https://assets.data.world/ic
ons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-120x120.png"><link rel="apple-touch-icon" sizes="144x144" href="http
s://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-144x144.png"><link rel="apple-touch-icon" sizes
="152x152" href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-152x152.png"><link rel="ap
ple-touch-icon" sizes="180x180" href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-touch-icon-180x1
80.png"><meta name="apple-mobile-web-app-capable" content="yes"><meta name="apple-mobile-web-app-status-bar-style" content="b
```

```
In [34]: british2_df.count()
```

Hacer cambios al data frame (columnas o renglones)

4. Hacerle cambios al Data Frame en Spark

```
In [ ]: Quitando los duplicados
```

```
In [36]: sub_nd_british2_df = british2_df.dropDuplicates()
```

```
In [37]: sub_nd_british2_df.describe().show()
```

```
count|
```

```
0|
```

```
0|    0|    0|    0|    0|    0|
```

```
0|        0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|            0|                0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

```
0|                0|    0|    0|    0|    0|    0|    0|    0|    0|    0|
```

Guarda el nuevo DF en HDFS y en el Bucket

5. Guarda el nuevo DF en HDFS y en el Bucket


```
In [39]: sub_nd_british2_df.write.save("/tmp/dcd/pyspark/british2/", format='csv', header=True)
```

```
In [ ]: sub_nd_british2_df.write.save("gs://dcd05omg-bucket/dcd/pyspark/british2/", format='csv', header=True)
```

```
In [ ]:
```


Agregar evidencia de que el JOB se ejecuto en el cluster

carbide-trees-398419 > cluster-sir-dcd05 [Sign out](#)

 **History Server**

Event log directory: gs://dataproc-temp-us-central1-351251296203-ufqeslfe/782eeaa4-0df7-4fa9-be40-060c18582916/spark-job-history

Last updated: 2023-09-22 02:55:01

Client local time zone: America/Mexico_City

Search:

Version	App ID	App Name	Driver Host	Started	Completed	Duration	Spark User	Last Updated	Event Log
3.3.2	application_1695356867203_0001	PySparkShell	cluster-sir-dcd05-m.us-central1-a.c.carbide-trees-398419.internal	2023-09-22 00:10:26	2023-09-22 00:46:43	36 min	root	2023-09-22 00:46:43	Download
3.3.2	application_1694277344266_0004	PySparkShell	cluster-sir-dcd05-m.us-central1-a.c.carbide-trees-398419.internal	2023-09-09 13:46:09	2023-09-09 13:46:41	32 s	sergio_ibarra1795	2023-09-09 13:46:42	Download
3.3.2	local-1694286223106	datosmasivos	cluster-sir-dcd05-m.us-central1-a.c.carbide-trees-398419.internal	2023-09-09 13:03:41	2023-09-09 13:13:08	9.5 min	sergio_ibarra1795	2023-09-09 13:13:09	Download

Showing 1 to 3 of 3 entries
[Show incomplete applications](#)

Mod13_Uso_Pyspark_tarea_sesion1

September 22, 2023

0.1 1. Descargar archivo referente a un tema en particular

0.1.1 Intentamos bajar la data de los archivos de British election study

Intentamos bajar la data de los archivos de British election study Part3

```
[1]: ! wget https://data.world/wwymak/british-election-study-part-3-csv-files/
      ↪ workspace/file?filename=BES2017_W13_Panel_v1.0-3.csv

--2023-09-22 06:47:36-- https://data.world/wwymak/british-election-study-
part-3-csv-files/workspace/file?filename=BES2017_W13_Panel_v1.0-3.csv
Resolving data.world (data.world)... 151.101.66.133, 151.101.2.133,
151.101.130.133, ...
Connecting to data.world (data.world)|151.101.66.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 32598 (32K) [text/html]
Saving to: 'file?filename=BES2017_W13_Panel_v1.0-3.csv.1'

file?filename=BES20 100%[=====>] 31.83K --.-KB/s in 0.003s

2023-09-22 06:47:37 (9.09 MB/s) - 'file?filename=BES2017_W13_Panel_v1.0-3.csv.1'
saved [32598/32598]
```

Borramos data que habíamos cargado anteriormente para solo tener archivos de British election study

```
[2]: import os

# Get the path to the CSV file.
csv_file_path = "amba-censo-csv.csv"

# Check if the CSV file exists.
if os.path.exists(csv_file_path):
    # Delete the CSV file.
    os.remove(csv_file_path)
```

Intentamos bajar la data de los archivos de British election study Part4

```
[3]: ! wget https://data.world/wwymak/british-election-study-part-3-csv-files/
      ↪ workspace/file?filename=BES2017_W13_Panel_v1.0-4.csv
```

```
--2023-09-22 06:47:59-- https://data.world/wwymak/british-election-study-
part-3-csv-files/workspace/file?filename=BES2017_W13_Panel_v1.0-4.csv
Resolving data.world (data.world)... 151.101.66.133, 151.101.2.133,
151.101.130.133, ...
Connecting to data.world (data.world)|151.101.66.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 32598 (32K) [text/html]
Saving to: 'file?filename=BES2017_W13_Panel_v1.0-4.csv.1'

file?filename=BES20 100%[=====>] 31.83K --.-KB/s in 0.003s

2023-09-22 06:47:59 (9.38 MB/s) - 'file?filename=BES2017_W13_Panel_v1.0-4.csv.1'
saved [32598/32598]
```

Intentamos bajar la data de los archivos de British election study Part1

```
[4]: ! wget https://data.world/wwymak/british-election-study-csv-files

--2023-09-22 06:48:07-- https://data.world/wwymak/british-election-study-csv-
files
Resolving data.world (data.world)... 151.101.194.133, 151.101.130.133,
151.101.2.133, ...
Connecting to data.world (data.world)|151.101.194.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 39212 (38K) [text/html]
Saving to: 'british-election-study-csv-files.1'

british-election-st 100%[=====>] 38.29K --.-KB/s in 0.004s

2023-09-22 06:48:08 (8.75 MB/s) - 'british-election-study-csv-files.1' saved
[39212/39212]
```

Intentamos bajar la data de los archivos de British election study Part2

```
[5]: ! wget https://data.world/wwymak/british-election-study-part-2-csv-files

--2023-09-22 06:48:11-- https://data.world/wwymak/british-election-study-
part-2-csv-files
Resolving data.world (data.world)... 151.101.194.133, 151.101.130.133,
151.101.2.133, ...
Connecting to data.world (data.world)|151.101.194.133|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 39606 (39K) [text/html]
Saving to: 'british-election-study-part-2-csv-files'

british-election-st 100%[=====>] 38.68K --.-KB/s in 0.004s

2023-09-22 06:48:11 (8.56 MB/s) - 'british-election-study-part-2-csv-files'
```

saved [39606/39606]

0.2 2. Cargar los archivos al cluster de HADOOP

Creación el directorio en el HDFS donde se almacenarán los archivos de csv de British election

```
[6]: ! hdfs dfs -mkdir -p /tmp/dcd/british/
```

```
[8]: ! hdfs dfs -ls -R /tmp/dcd
```

```
drwxr-xr-x  - root                hadoop              0 2023-09-22 06:51
/tmp/dcd/british
drwxr-xr-x  - root                hadoop              0 2023-09-09 19:52
/tmp/dcd/ecobici
-rw-r--r--   2 root                hadoop            41926344 2023-09-09 19:52
/tmp/dcd/ecobici/2023-01.csv
-rw-r--r--   2 root                hadoop            47532884 2023-09-09 19:52
/tmp/dcd/ecobici/2023-02.csv
-rw-r--r--   2 root                hadoop            58449187 2023-09-09 19:52
/tmp/dcd/ecobici/2023-03.csv
-rw-r--r--   2 root                hadoop            58119473 2023-09-09 19:52
/tmp/dcd/ecobici/2023-04.csv
-rw-r--r--   2 root                hadoop            67532130 2023-09-09 19:52
/tmp/dcd/ecobici/2023-05.csv
-rw-r--r--   2 root                hadoop            72151312 2023-09-09 19:52
/tmp/dcd/ecobici/2023_06.csv
-rw-r--r--   2 root                hadoop            71416339 2023-09-09 19:52
/tmp/dcd/ecobici/2023_07.csv
-rw-r--r--   2 root                hadoop            79235355 2023-09-09 19:52
/tmp/dcd/ecobici/2023_08.csv
drwxr-xr-x  - sergio_ibarra1795 hadoop              0 2023-09-09 18:39
/tmp/dcd/sirilo
-rw-r--r--   2 sergio_ibarra1795 hadoop          143562785 2023-09-09 18:39
/tmp/dcd/sirilo/linea-mujeres-cdmx.csv
drwxr-xr-x  - sergio_ibarra1795 hadoop              0 2023-09-09 19:33
/tmp/dcd/wordcount
drwxr-xr-x  - sergio_ibarra1795 hadoop              0 2023-09-09 18:03
/tmp/dcd/wordcount/input
-rw-r--r--   2 sergio_ibarra1795 hadoop           2226045 2023-09-09 18:03
/tmp/dcd/wordcount/input/texto.txt
drwxr-xr-x  - sergio_ibarra1795 hadoop              0 2023-09-09 18:09
/tmp/dcd/wordcount/output
-rw-r--r--   2 sergio_ibarra1795 hadoop              0 2023-09-09 18:09
/tmp/dcd/wordcount/output/_SUCCESS
-rw-r--r--   2 sergio_ibarra1795 hadoop           148653 2023-09-09 18:09
/tmp/dcd/wordcount/output/part-r-00000
-rw-r--r--   2 sergio_ibarra1795 hadoop           149555 2023-09-09 18:09
```

```

/tmp/dcd/wordcount/output/part-r-00001
-rw-r--r-- 2 sergio_ibarra1795 hadoop 148719 2023-09-09 18:09
/tmp/dcd/wordcount/output/part-r-00002
drwxr-xr-x - sergio_ibarra1795 hadoop 0 2023-09-09 19:33
/tmp/dcd/wordcount/pyspark
-rw-r--r-- 2 sergio_ibarra1795 hadoop 0 2023-09-09 19:33
/tmp/dcd/wordcount/pyspark/_SUCCESS
-rw-r--r-- 2 sergio_ibarra1795 hadoop 324685 2023-09-09 19:33
/tmp/dcd/wordcount/pyspark/part-00000
-rw-r--r-- 2 sergio_ibarra1795 hadoop 321210 2023-09-09 19:33
/tmp/dcd/wordcount/pyspark/part-00001

```

Se suben todos los archivos de tipo csv que se han descargado dentro del directorio de british

```
[7]: ! hdfs dfs -put *.csv /tmp/dcd/british/
```

```

put: `file': No such file or directory
put: `file': No such file or directory

```

```

[9]: import os

# Get the current working directory.
cwd = os.getcwd()

# Print the current working directory.
print(cwd)

```

/

```
[ ]: Se tuvo que renombrar los archivos de csv de British election para lograr un
    ↪ mejor manejo
```

```

[10]: import os

# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
file_path = os.path.join(cwd, "file?filename=BES2017_W13_Panel_v1.0-3.csv.1")

# Rename the file.
os.rename(file_path, "bes2017_part3.csv")

```

```
[ ]: Y se intenta subir de nuevo el archivo bes2017_part3.csv a la carpeta british
    ↪ dentro de mi cluster de Hadoop
```

```
[11]: ! hdfs dfs -put bes2017_part3.csv /tmp/dcd/british/
```

[]: Se verifica que el archivo de nombre bes2017_part3.csv se haya subido
↪correctamente a /tmp/dcd/british/

[12]: ! hdfs dfs -ls -R /tmp/dcd/british

```
-rw-r--r--  2 root hadoop      32598 2023-09-22 07:09
/tmp/dcd/british/bes2017_part3.csv
```

[]: Se repite el proceso de renombrar y subir archivos a la carpeta british

[13]: `import os`

```
# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
file_path = os.path.join(cwd, "file?filename=BES2017_W13_Panel_v1.0-4.csv.1")

# Rename the file.
os.rename(file_path, "bes2017_part4.csv")
```

[]: Y se intenta subir de nuevo el archivo bes2017_part4.csv a la carpeta british
↪dentro de mi cluster de Hadoop

[14]: ! hdfs dfs -put bes2017_part4.csv /tmp/dcd/british/

[16]: `import os`

```
# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
file_path = os.path.join(cwd, "british-election-study-csv-files.1")

# Rename the file.
os.rename(file_path, "bes2017_part1.csv")
```

[]: Y se intenta subir de nuevo el archivo bes2017_part1.csv a la carpeta british
↪dentro de mi cluster de Hadoop

[17]: ! hdfs dfs -put bes2017_part1.csv /tmp/dcd/british/

[18]: `import os`

```
# Get the current working directory.
cwd = os.getcwd()

# Get the path to the downloaded file.
```

```
file_path = os.path.join(cwd, "british-election-study-part-2-csv-files")

# Rename the file.
os.rename(file_path, "bes2017_part2.csv")
```

[]: Y se intenta subir de nuevo el archivo bes2017_part1.csv a la carpeta british_ dentro de mi cluster de Hadoop

[19]: ! hdfs dfs -put bes2017_part2.csv /tmp/dcd/british/

[24]: ! hdfs dfs -ls -R /tmp/dcd/british

```
-rw-r--r--  2 root hadoop      39212 2023-09-22 07:14
/tmp/dcd/british/bes2017_part1.csv
-rw-r--r--  2 root hadoop      39606 2023-09-22 07:15
/tmp/dcd/british/bes2017_part2.csv
-rw-r--r--  2 root hadoop      32598 2023-09-22 07:09
/tmp/dcd/british/bes2017_part3.csv
-rw-r--r--  2 root hadoop      32598 2023-09-22 07:13
/tmp/dcd/british/bes2017_part4.csv
```

[27]: `import os`

```
# Get the path to the file you want to delete.
file_path3 = "/tmp/dcd/british/bes2017_part3.csv"

# Delete the file.
os.remove(file_path3)
```

```
-----
FileNotFoundError                                Traceback (most recent call last)
/tmp/ipykernel_20152/1778995922.py in <cell line: 7>()
      5
      6 # Delete the file.
----> 7 os.remove(file_path3)

FileNotFoundError: [Errno 2] No such file or directory: '/tmp/dcd/british/
bes2017_part3.csv'
```

[]:

[28]: ! hdfs dfs -mkdir -p /tmp/dcd/british2/

[29]: ! hdfs dfs -put bes2017_part3.csv /tmp/dcd/british2/

[30]: ! hdfs dfs -put bes2017_part4.csv /tmp/dcd/british2/

```
! hdfs dfs -ls -R /tmp/dcd/british2
```

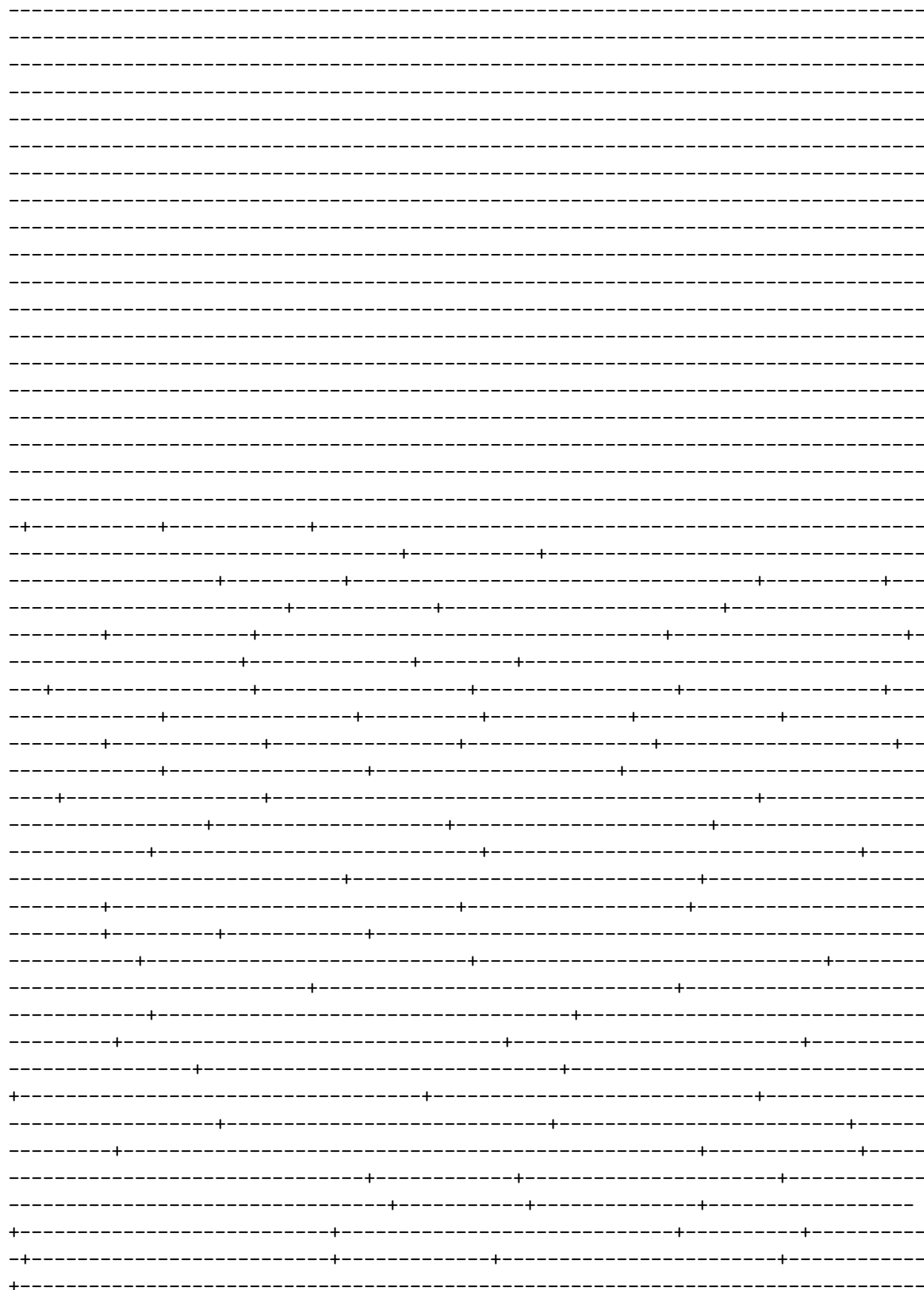
```
-rw-r--r--    2 root  hadoop      32598 2023-09-22 08:39
/tmp/dcd/british2/bes2017_part3.csv
-rw-r--r--    2 root  hadoop      32598 2023-09-22 08:39
/tmp/dcd/british2/bes2017_part4.csv
```

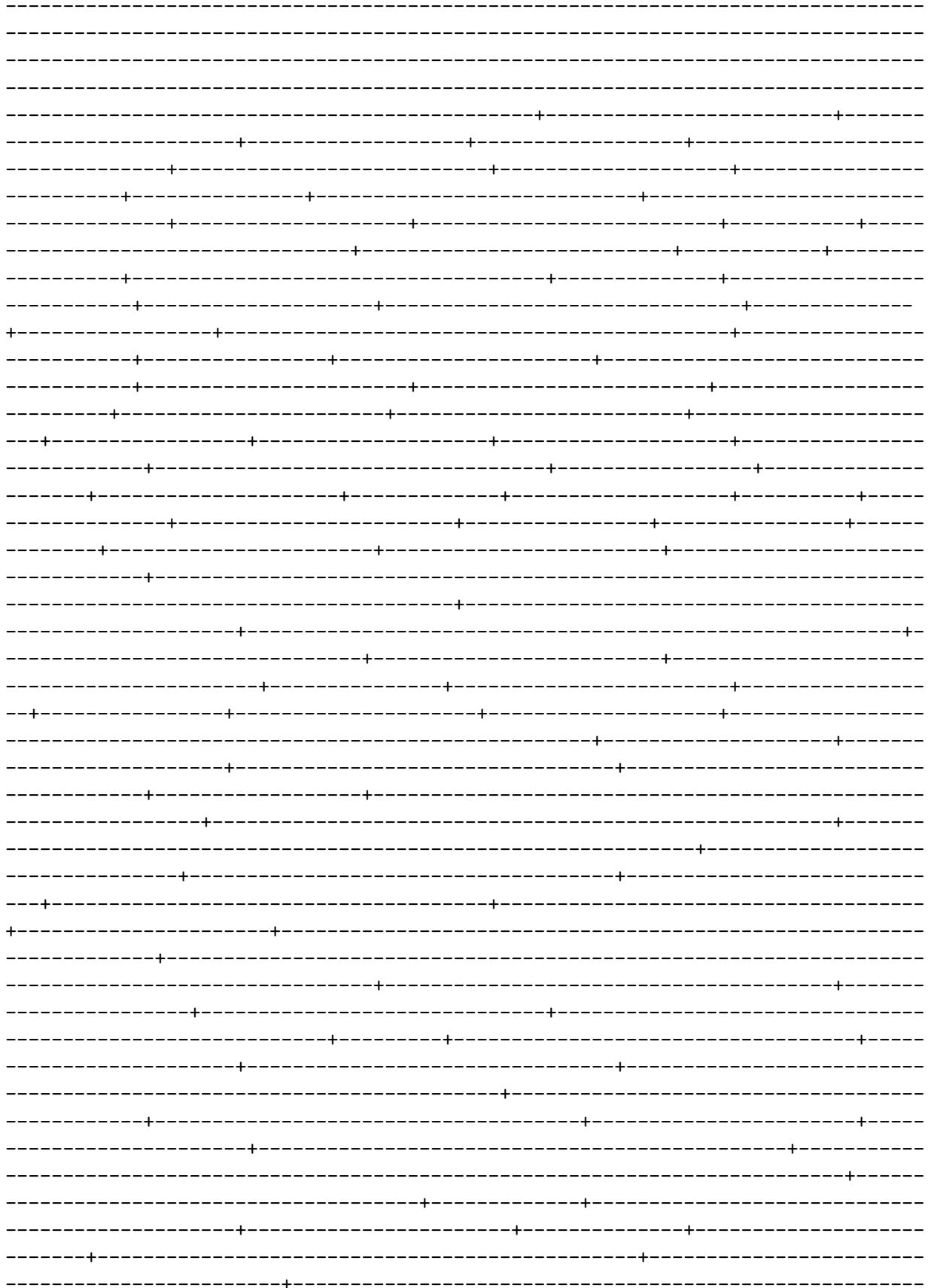
0.3 3. Crear un Data Frame en Spark

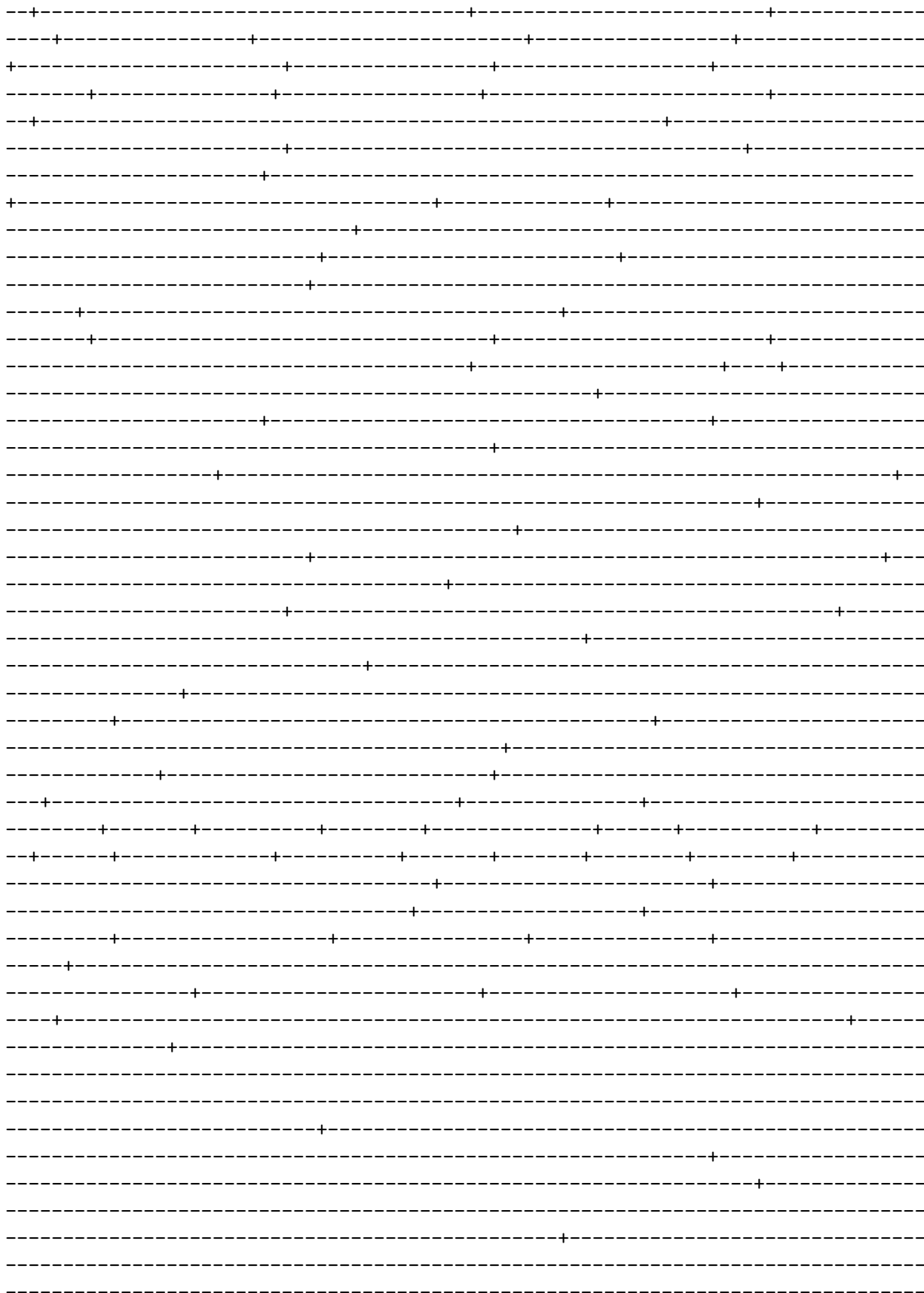
```
british2_df = spark.read.csv('hdfs:///tmp/dcd/british2/', header=True, inferSchema=True)
```

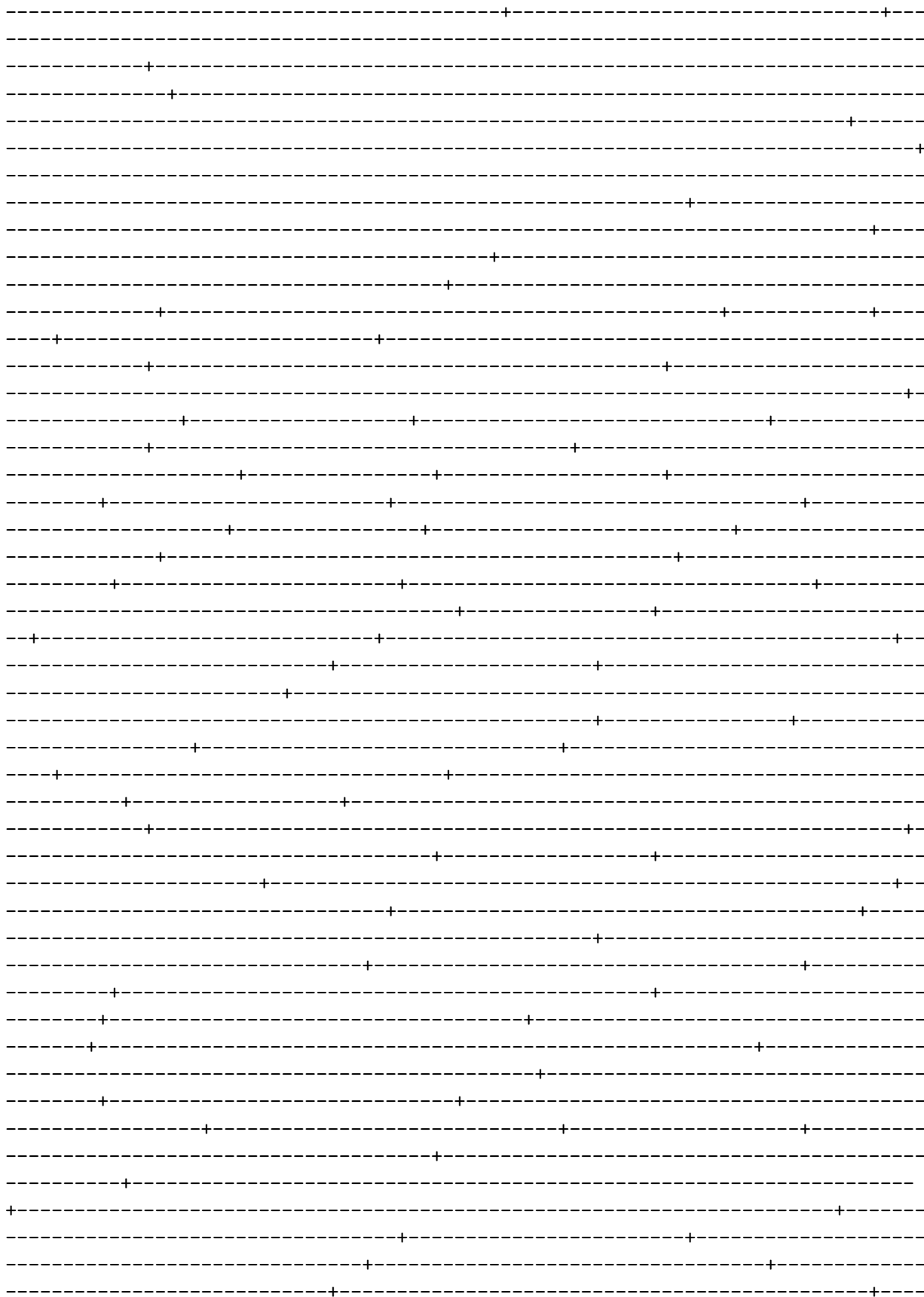
```
british2_df.show(10)
```

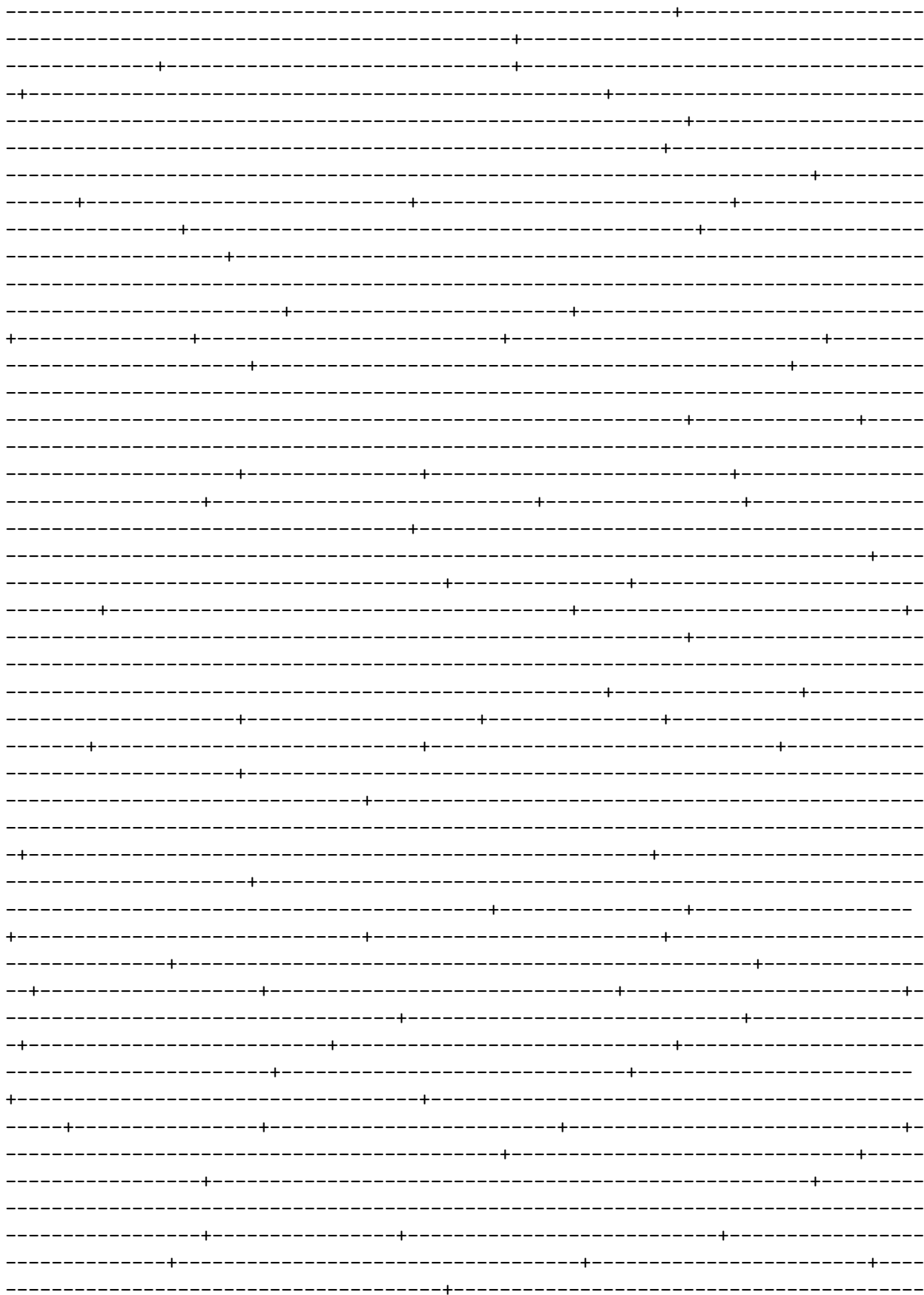
[illegible]

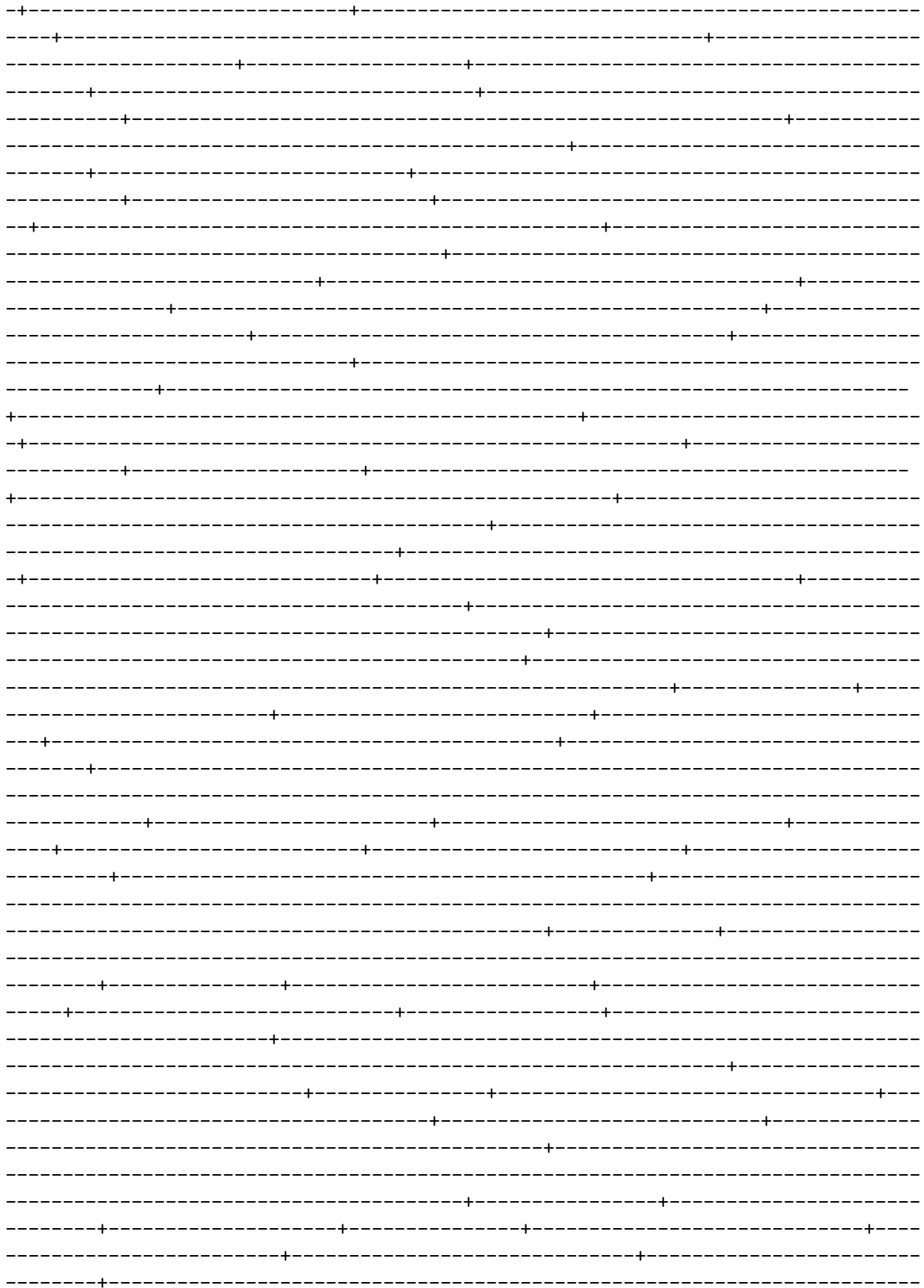


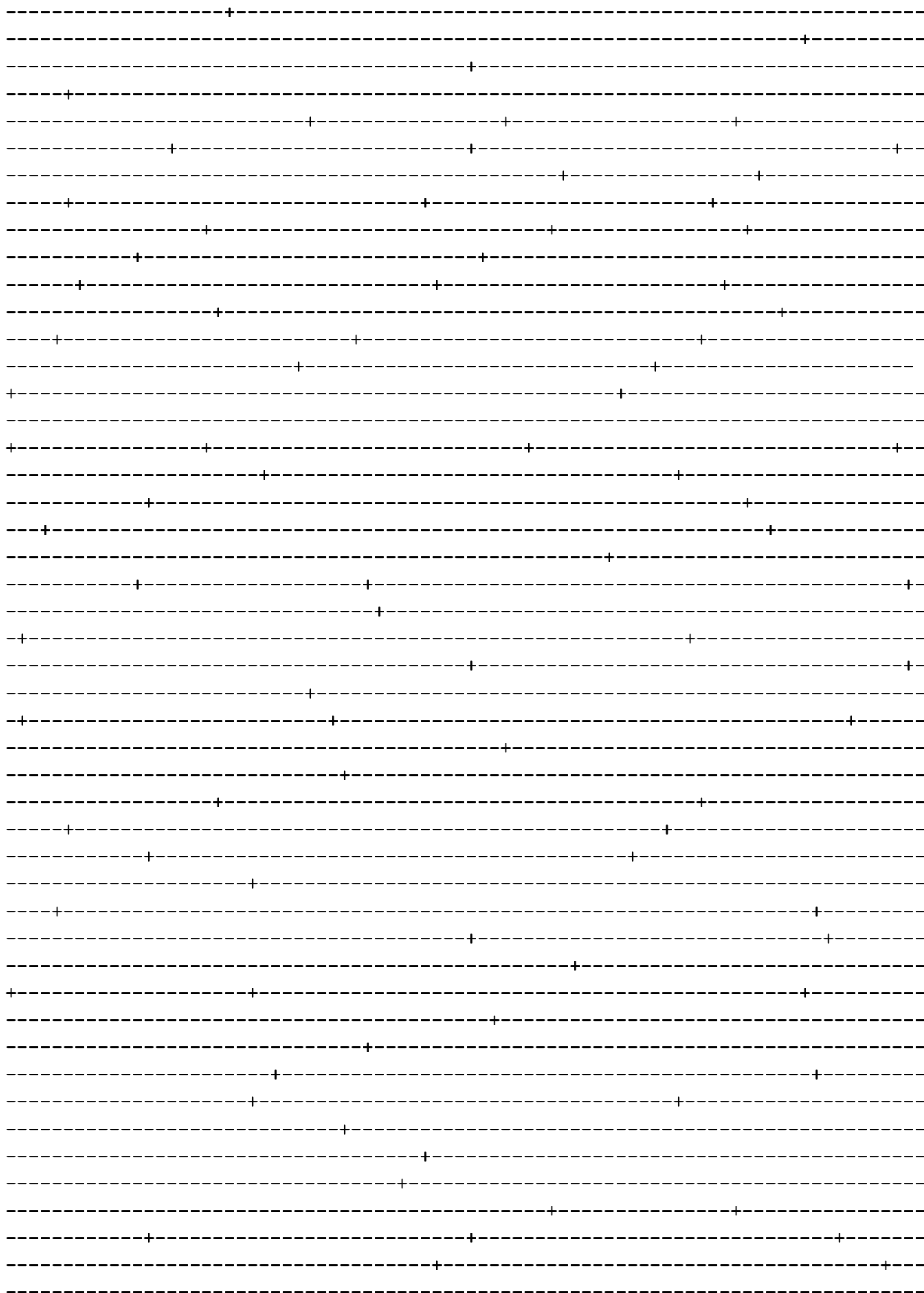


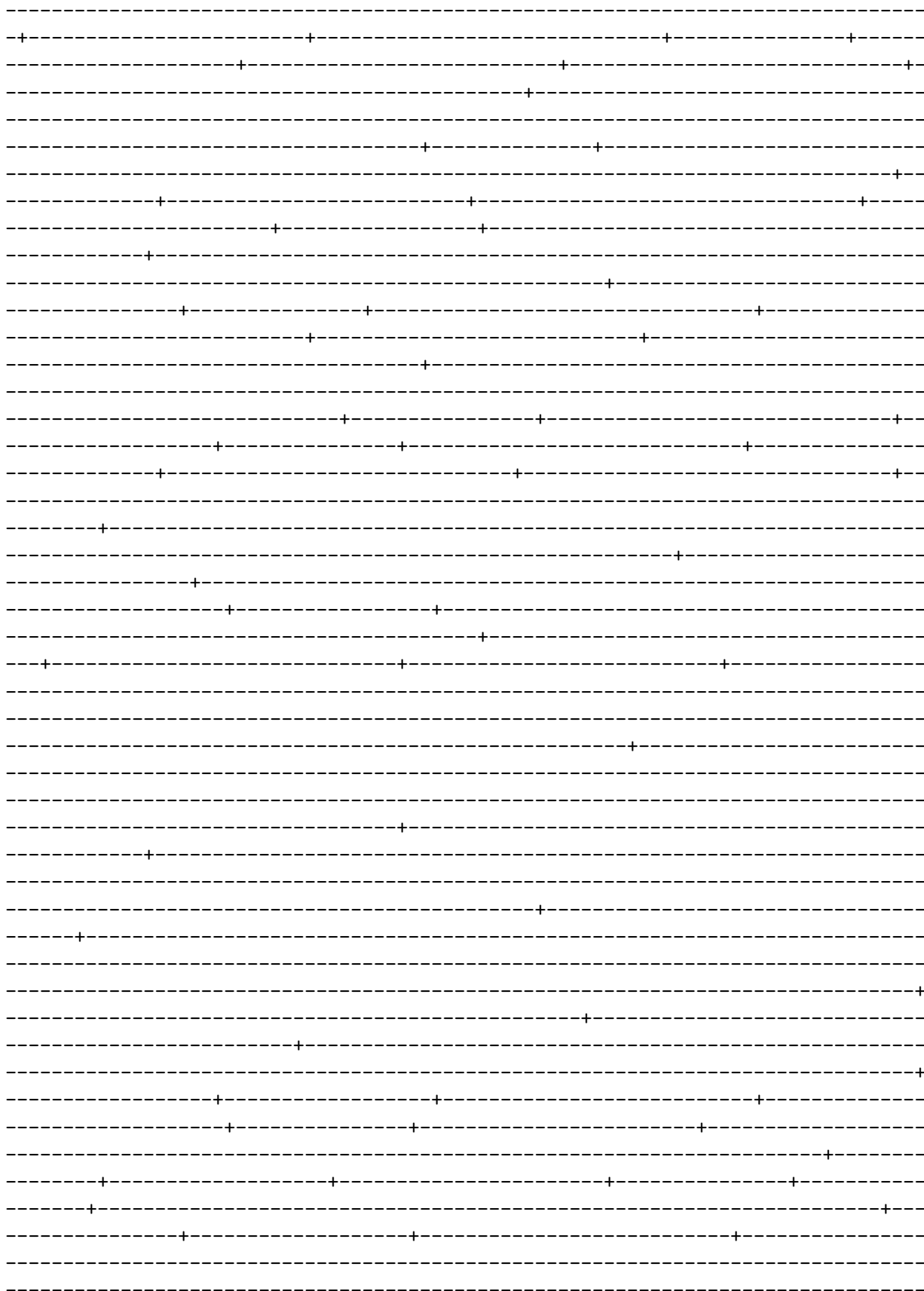













```

-----+
|<!doctype html><html lang="en"><head> <title data-react-
helmet="true">data.world</title>    <meta charset="utf-8"><meta name="viewport"
content="width=device-width|initial-scale=1"><meta name="referrer"
content="origin-when-cross-origin"><meta name="og:image"
content="https://d2ogkqlrg66kh0.cloudfront.net/site-resources/images/fb-image-
share.3a87a6eb.png"><meta name="og:site_name" content="data.world"><meta
name="twitter:card" content="summary"><meta name="twitter:site"
content="@datadotworld"><meta property="fb:app_id"
content="776416075825233"><link rel="stylesheet" href="https://fonts.googleapis.
com/css2?family=Inconsolata:wght@400;700&family=Lato:ital|wght@0|300;0|400;0|700
;1|300;1|400;1|700&display=swap"><link rel="search" href="/opensearch.xml"
type="application/opensearchdescription+xml" title="data.world"> <link
rel="apple-touch-icon" sizes="57x57"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-57x57.png"><link rel="apple-touch-icon" sizes="60x60"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-60x60.png"><link rel="apple-touch-icon" sizes="72x72"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-72x72.png"><link rel="apple-touch-icon" sizes="76x76"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-76x76.png"><link rel="apple-touch-icon" sizes="114x114"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-114x114.png"><link rel="apple-touch-icon" sizes="120x120"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-120x120.png"><link rel="apple-touch-icon" sizes="144x144"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-144x144.png"><link rel="apple-touch-icon" sizes="152x152"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-152x152.png"><link rel="apple-touch-icon" sizes="180x180"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-180x180.png"><meta name="apple-mobile-web-app-capable"
content="yes"><meta name="apple-mobile-web-app-status-bar-style" content="black-
translucent"><meta name="apple-mobile-web-app-title" content="data.world"><meta
name="mobile-web-app-capable" content="yes"><meta name="theme-color"
content="#fff"><meta name="application-name" content="data.world"><link
rel="icon" type="image/png" sizes="32x32" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-32x32.png"><link rel="icon"
type="image/png" sizes="16x16" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-16x16.png"><link rel="shortcut icon" hr

```

```

ef="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon.ico"><link
href="https://assets.data.world/css/vendors~app.bf2ab82ca4e8e237d22a.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors~vendors.e418760f21d557e2e0e2.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors.5b2975e4af1441b468c5.css"
rel="stylesheet"><link
href="https://assets.data.world/css/app.1915b3c34db820707a63.css"
rel="stylesheet"></head><body ><div id="root"><div class="page-container
CoreLayout__embed__1fxWq"><div class="view-container"><div
class="LoadingAnimation__loaderOverlay__2zSul
LoadingAnimation__default__KhFF6" data-dw="LoadingAnimation"><div
class="LoadingAnimation__loaderContainer__2arLY"><div class="LoadingAnimation__
spinner__2qQFp">Loading...</div></div></div></div></div></div></div><script
type="application/json" id="init-data">{"catalog":{"loadedResources":{}|"hoots":
{}|"clients":{}|"currentUser":{"profile":{"abTests":{"buckets":{"coreLayoutRefa
ctor":{"name":"inactive"|"value":-1|"loadInstanceConfig":{"name":"load-
instance-config"|"value":1|"sqlAlgebra":{"name":"preferLegacy"|"value":2|"sqlQ
uery":{"name":"presto"|"value":4}}|"accountStanding":"good"|"activeSubscription
id":"","|agentid":"","|authorizationLevels":{"org":"READ"|"allCatalogs":"NONE"|"da
taLibrary":"NONE"|"avatarUrl":"","|bio":"","|catalogStatistics":{"analysesCount":
0|"datasetsCount":0|"dataTypesCount":0|"glossaryCount":0|"projectsCount":0|"tabl
esCount":0|"displayName":"","|email":"","|created":"","|company":"","|emailVerified
:true|"location":"","|numFollowers":0|"numFollowing":0|"numOrganizations":0|"org
Mode":false|"orgDetails":null|"orgPreferences":null|"relationship":{"member":fal
se|"followed":false|"staticGrants":{"analysisHubEnabled":false|"businessGlossar
yEnabled":false|"hootsEnabled":false|"lineageEnabled":false|"metadataCatalogEnab
led":false|"nativeLineageEnabled":false|"personalDatasetsDisabled":false|"preven
tGuestAuthorizations":false|"preventGuestUserInvite":false|"preventOpenDatasets"
:false|"preventSignedShareUrls":false|"samlAllowed":false|"thirdPartyAiEnabled":
false|"tags":[]|"updated":"","|userCapabilities":{"canBulkEditAllMetadataResourc
es":false|"canCreateCollections":false|"canCreateDataResources":false|"canCreate
MetadataResources":false|"canDebugMetadataProfile":false|"canEditAllDatasetMetad
ata":false|"canManageAllCollectionAccess":false|"canManageAllCollectionHierarchy
":false|"canManageAllDataResources":false|"canManageAllHoots":false|"canManageOr
gAccess":false|"canManageOrgAutomations":false|"canManageOrgConnections":false|"
canManageOrgConnectionTasks":false|"canManageOrgSettings":false|"canManageOrgSub
scription":false|"canShowOrgAdminLabel":false|"hasOrgMembership":false|"visibil
ity":"","|visitorid":"109156be-c4fb-41ea-b1b4-
efe1671c5808"|"website":"","|authorizedAccounts":{"state":"idle"|"data":null}|"a
uthorizedClients":[]|"customSearch":{"default":{"customFacets":[]|"byOrg":{}}|"
dismissed":[]|"isLoggedIn":false|"needsAnalyticsAlias":false|"notifications":{"r
esults":[]|"unread":0|"unseen":0|"onboardThisSession":false|"requests":[]|"requ
estsFetching":false|"tasks":null|"token":"eyJhbGciOiJIUzUxMiJ9.eyJzdWIiOiJwcm9kL
XVzZXItY2xpZW50IiwiaXNzIjoieYXV0aG9yaXR5OmRhZGFkb3R3b3JsZDo6MEJGMEVCRDMtMkRfNy00O
UZCLUi4NOitQUQwMzA1NOJCODlFIiwiaWF0IjoxNjk1MzMxMjM1LCJyb2x1IjpbXSwiZXhwIjoxNjk1N
DE3NjM1LCJ2aXNpdG9yaWQiOiIxMDkxNTZiZS1jNGZiLTQxZWVtYjFiNC1lZmUxNjcxYzU4MDgiLCJnZ

```

W5lcmFsLXB1cnBvc2UiOnRydWUsImF1dGhvcm10eWlkcyI6WyJkYXRhZG90d29ybGQiXX0.nFFhdkNCoIGQuVLekewk_RfKKtP4WTwb1raXlseSfGgyGhI6y5x1f5ZFi80mOLpreLK7ySuqAHA88MFwOYP8AA"|"undismissedTourCount":2|"dataset":{"hasError":false|"loadedDatasets":{"usersDatasets":{"loadedCurrentUsersDatasets":false|"usersDatasetsTotalCount":0|"customDataTypes":{"workspacePreferences":{"discussion":{"events":{"status":"closed"}}|"integration":{"activeIntegrations":[]|"userJdbcSources":[]|"supportedJdbcSources":[]|"invite":{"linkedOnboardModal":{"show":false|"loginModal":{"show":false|"queries":{"relationships":{"requestInfo":{"pathsMarkedAs404":[]|"metadata":{"requests":{"hasError":false|"loadedResources":{"allPendingRequests":{"count":0|"records":null|"hydrations":{"}}|"routing":{"locationBeforeTransitions":null|"signupModal":{"show":false|"treatment":null|"treatmentProps":null|"siteStatus":{"uploadsOperational":true|"queryOperational":true|"websiteOperational":true|"downloadsOperational":true|"shouldAskReload":false|"shouldForceReload":false|"connectionError":false|"incidentName":"","incidentMessage":"","pageStatus":"none"}}|"treatments":{"isSite":false|"activeTreatments":["showVersions|showWorkspaceTabs|showEntityLinkOptions|canCreateOpenDatasets|canCreateOrgs|showConversionPages|showSignup|showSocialLogin"]|"user":{"hasError":false|"loadedUsers":{"userFollows":{"userLikes":{"userRecentComments":{"groupAuthorizations":{"featuredDatasets":{"loaded":false|"resources":[]}}}}</script><script type="application/json" id="init-config">{"adwords_pixel_id":"869213640"|"analytics":{"segment_id":"IXhtZmPohuGM60VHk59cdSbWqWQBB7aR"|"segment_endpoint":"https:\u002F\u002Fapi-el.data.world"|"analytics_log_to_console":false|"base_cookie_domain":null|"contact_admin_form_config":[{"type":"input"|"label":"Email"|"lookupProperty":"email"|"readOnly":true|162|"required":true}]{type:"textarea"164|"label":"Message"165|"placeholder":"Add a message (i.e details that need to be reviewed)"|"required":true}}167|"csrf_cookie_name":"_csrf"|"csrf_header_name":"x-csrf-token"|"data_provider_agents":["johnsnowlabs"|"associatedpress"]|"download_content_url":"https:\u002F\u002Fdownload.dwcontent.com"|"download_url":"https:\u002F\u002Fdownload.data.world"|"dw_tableau_wdc_host_domain":"https:\u002F\u002Ftableau.data.world"|"deployed_environment_full":"prod"|"facebook_pixel_id":"261984994177057"|"fb_app_id":"776416075825233"|"google_choser_app_id":"956984791354"|"help_cards":[{"id":"getting_started"|"header":"Get Started"|"body":"New to data.world? Start here to find introductory articles."|"link":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"|"imageSrc":"getting-started-lightglow"}]{id:"join_discussion"|"header":"Join the discussion"|"body":"Join the general discussion or start your own topic"|"link":"","imageSrc":"discussions-lightglow"}]{id:"see_documentation"|"header":"Visit the Help Center"|"body":"Find general platform documentation and technical tutorials."|"link":"https:\u002F\u002Fdocs.data.world\u002F"|"imageSrc":"documentation-lightglow"}]{id:"contact_support"|"header":"Contact support"|"body":"Report an issue or make a suggestion."|"link":"https:\u002F\u002Fhelp.data.world\u002Fhc\u002Fen-us\u002Frequests\u002Fnew"|"imageSrc":"support-lightglow_sales-guy"}}|"hoot_url":"","iframely_key":"e4239223dd4ad21531b3a044840d5898"|"is_monitored_env":true|"is_vpc":false|"max_upload_size":524288000|"media_upload_domain":"mediauploads.data.world"|"media_url":"https:\u002F\u002Fmedia.data.world"|"oauth_authorize_url":

"https:\u002F\u002Fauth.data.world"|"oauth_providers":[{"name":"facebook"|"display_name":"Facebook"|"active":true}209|{"name":"github"|"display_name":"GitHub"|"active":true}212|{"name":"google"|"display_name":"Google"|"active":true}215|{"name":"linkedin"|"display_name":"LinkedIn"|"active":false}|{"name":"twitter"|"display_name":"Twitter"|"active":true}]|"openAccessRoutePrefixes":["\u002Fplaindeale rprojects"|\u002Fd1gi\u002Fmissing-fb-posts-w-share- stats|\u002Ftravelwarnings\u002Ftravel- danger|\u002Ffootprint\u002Fnfa-2018-edition|\u002Fpolymathic\u002Fcasualties- of-the-syrian-civil- war|\u002Faurielle\u002Ffortune-500-2017|\u002Fnjdotcom|\u002Fian\u002Fusa-baby- names-national-and-state-by-state|\u002Fowentemple\u002Ftext-and-content- features-of-most-persuasive-ted-talks|\u002Fschnee\u002Ftangled|\u002Fchhs\u002F fc544658-35c5-4be0-af20-fc703bc57c13|\u002Fnz-stats- nz\u002F4e5ebed0-4279-4292-a85f-33c9f2a93f8d|\u002Fshad\u002Fcovid-19-time- series-data|\u002Fmarkmarkoh\u002Fcoronavirus-data|\u002Fdallemang\u002Fcovid- 19-data|\u002Fpolicinginamerica|\u002Findustryresearch\u002Fdata-engineer- survey-2021|\u002Fembed\u002Fswwo|swwo|\u002Fassociatedpress\u002F2018-election- results-precinct-level|\u002Faaron-kessler-ap\u002Fcongressional- earmarks|\u002Fassociatedpress\u002Fap- votecast|\u002Fassociatedpress\u002Fbolsoa-familia-aid-to-brazils- poor|\u002Fassociatedpress\u002Fcdc-covid-19-vaccine- tracker|\u002Fassociatedpress\u002Fcovid-impact-survey-public- data|\u002Fassociatedpress\u002Fcovid-tracking-project-testing-in- states|\u002Fassociatedpress\u002Fap-planner-covid-reopening- data|\u002Fassociatedpress\u002Ffema-distribution-of-ppe-to- states|\u002Fassociatedpress\u002Fforeign-exchange- rates|\u002Fassociatedpress\u002Fus-dams- public|\u002Fassociatedpress\u002Fjohns-hopkins-coronavirus-case- tracker|\u002Fassociatedpress\u002Fmass-killings- public|\u002Fassociatedpress\u002Fopioid- prescriptions-2010-2015|\u002Fassociatedpress\u002Fpublic-health-official- departures|\u002Fassociatedpress\u002Frefugee-admissions-to-us-end- fy-2018|\u002Fassociatedpress\u002Frental-assistance-from-coronavirus-relief- fund|\u002Fassociatedpress\u002Fstate-ppe- purchases|\u002Fassociatedpress\u002Fmarshall-project-covid-cases-in- prisons|\u002Fassociatedpress\u002Fvehicle-miles-traveled|\u002Frare\u002Ffish- surveys|"|\u002Frare\u002Flarval-dispersal-data"]|"preferred_oauth_provider":"goo gle"|"prod_env":true|"profile_interests":["education"|"finance"|"government"|"busines s"|"transportation"|"health"|"agriculture"|"environment"|"sports"|"entertainment"|"technology"|"science"|"culture"|"politics"|"crime"]|"query_url":"https:\u002F\u002Fquery.data.world"|"regional_cluster":"US"|"request_dataset_access_form_config":[{"type":"input"|"label":"Full name"|"lookupProperty":"displayName"284|"required":true285|"readO nly":true}|{"type":"email"|"label":"Contact email"|"placeholder":"Provide an email in the case a dataset administrator needs to reach out"289|"showOptionalFl ag":true}|{"type":"textarea"291|"label":"Message"292|"placeholder":"Share what you would like to do with this dataset"293|"required":true}]294|"requests":{"con tributionSuggestionInfoText":"You are suggesting changes to this \u003C%= resourceType %\u003E that will need to be approved by an admin of this

organization. Only you and admins will be able to see your changes until they're approved."|"pendingApprovalsText":"There \u003C%= pendingApprovalsNumber %\u003E from the community pending admin approval."|"pendingSuggestionsText":"Your suggested changes are visible on the Suggestions tab| but only you and this organization's admins can see them until they have been approved. The Overview tab displays only approved information."|"personalResourceSharedWithText":"This draft \u003C%= resourceType %\u003E is within your private collection and only visible to you. It is neither discoverable nor viewable by the data.world community. To open it up to others to view| publish it to an organization."|"transferOwnershipMessage":"By publishing this \u003C%= resourceType %\u003E to an organization| you are transferring ownership to the organization. Once your draft is published| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."|"transferOwnershipModalWarningBanner":"Once the organization has published your draft| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."|"sentry_dsn_client":"https:\u002F\u002Ff763f0cf14f04c75b4789a118ab1beb3@sentry.io\u002F1486857"|"server_origin":"https:\u002F\u002Fdata.world"|"sparql_host":"https:\u002F\u002Fpdw-sparqlproxy-us-east-1.prod.data.world"|"stripe_public_key":"pk_live_CPGU5gVFxlkY6pabDXladsMR"|"swagger_use_case_label_override_org":["oa-test"|engie-impact|"e-ddw"]|"tableau_api_version":"2.8"|"ui":{"orgAdmin":{"resourceLinks":[{"key":"getting-started"|"linkDisplayName":"Getting started guide"315|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"316|"newTab":true317|"svgIcon":"meta"318|"verticallyAlignIcon":true}319|{"key":"documentation"320|"linkDisplayName":"Documentation"321|"url":"https:\u002F\u002Fdocs.data.world\u002F"322|"newTab":true323|"svgIcon":"meta"324|"verticallyAlignIcon":true}325|{"key":"integrations"326|"linkDisplayName":"Integrations"327|"url":"\u002Fintegrations"328|"newTab":true329|"svgIcon":"integration"330|"verticallyAlignIcon":true331|"showIfAnyDismissed":["onboard-checklist"332|"onboard-checklist-admin"]}}333|{"key":"new-features"334|"linkDisplayName":"New features"335|"url":"https:\u002F\u002Fwhatsnew.data.world"336|"newTab":true337|"svgIcon":"magic-wand"338|"verticallyAlignIcon":true339|"showIfAnyDismissed":["onboard-checklist"340|"onboard-checklist-admin"]}}341|{"key":"enterprise"342|"linkDisplayName":"data.world for the enterprise"343|"url":"https:\u002F\u002Fdata.world\u002Fsched-a-demo\u002F?src=inapp_data.world-for-the-enterprise"344|"newTab":true345|"svgIcon":"organization"346|"verticallyAlignIcon":true}}347|"onboardChecklist":{"key":"site"348|"items":[{"id":"createAccount"349|"analyticsId":"onboard_checklist.create_account"350|"complete":true351|"formattedMessage":{"id":"onboard_checklist.create_account"352|"description":"Text for onboard checklist create account link"353|"defaultMessage":"Create account"}}354|{"id":"inviteTeam"|"analyticsId":"onboard_checklist.invite-team"|"href":"\u002F\u003C%- orgAgentid %\u003E?tab=members"|"dismissOnClick":"invite-team-link"|"completeIfDismissed":"invite-team-link"|"formattedMessage":{"id":"onboard_checklist.invite-

```

team|"description":"Text for onboard checklist invite team
link|"defaultMessage":"Invite your
team"}}|{"id":"connectData|"analyticsId":"onboard_checklist.connect-
data|"href":"\u002Fcreate-a-dataset|"dismissOnClick":"connect-data-
link|"completeIfDismissed":"connect-data-
link|"formattedMessage":{"id":"onboard_checklist.connect-
data|"description":"Text for onboard checklist connect data
link|"defaultMessage":"Connect data sources"}}|{"id":"exploreIntegrations"371|"
analyticsId":"onboard_checklist.explore-integrations"372|"href":"\u002Fintegrati
ons"373|"target":"_blank"374|"dismissOnClick":"explore-integrations-
link"375|"completeIfDismissed":"explore-integrations-
link"376|"formattedMessage":{"id":"onboard_checklist.explore-
integrations"377|"description":"Text for onboard checklist explore integrations
link"378|"defaultMessage":"Explore integrations"}}|{"id":"createProject"380|"ana
lyticsId":"onboard_checklist.create-project"381|"href":"\u002Fcreate-a-
project"382|"dismissOnClick":"create-project-
link"383|"completeIfDismissed":"create-project-
link"384|"formattedMessage":{"id":"onboard_checklist.create-
project"385|"description":"Text for onboard checklist create project
link"386|"defaultMessage":"Create your first
project"}}}|{"onboardTour":[{"show":true388|"header":"Welcome to
data.world!"389|"subhead":"The place for modern data
teamwork."390|"leftCol":"Organize data work so you and your team find what you
need&nbsp;faster."391|"middleCol":"Explore and create a shared understanding of
your&nbsp;data."392|"rightCol":"Share insights and invite more people to
participate in data&nbsp;teamwork."}393|{"show":true394|"imageHeaderKey":"search
"395|"imageKey":"collaborate"396|"imageAlt":"Start with Data"397|"header":"Start
with fantastic&nbsp;data"398|"subhead":"Browse thousands of free399| open data
resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from
almost anywhere into a single searchable&nbsp;library.\n\nMake it open400| share
with your team401| or keep it&nbsp;private."}402|{"show":true403|"imageHeaderKey
":"bulb"404|"imageKey":"dataProject"405|"imageAlt":"Data
Projects"406|"header":"Organize your data&nbsp;projects"407|"subhead":"Create
projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd
context to make it easier for others to discover408| understand409| and
contribute to your&nbsp;projects.\n\nImmediately query data without leaving your
&nbsp;projects."}410|{"show":true411|"imageHeaderKey":"code"412|"imageKey":"slid
eIntegrations"413|"imageAlt":"Integrations"414|"reverse":true415|"header":"Conne
ct the data tools you&nbsp;love"416|"subhead":"Automatically keep data.world
synced with your data&nbsp;sources.\n\nIntegrate your team's favorite tools417|
from analysis to&nbsp;visualization."}418|{"show":true419|"imageHeaderKey":"disc
ussions"420|"imageKey":"discussionsCollaboration"421|"imageAlt":"Contributors"42
2|"header":"Add more brainpower to&nbsp;your&nbsp;projects"423|"subhead":"Invite
your team to securely and effectively work&nbsp;together.\n\nManage fine-grained
permissions for each&nbsp;member.\n\nAsk questions424| add insights425| bookmark
top contributions426| and&nbsp;more."}427|{"show":true428|"imageHeaderKey":"insi
ght"429|"imageKey":"knowYourData"430|"imageAlt":"know your
data"431|"header":"Share your results"432|"subhead":"Share meaningful insights

```

and the data behind them. Distribute clean documented data for others to use in their own work. Help the right people discover the resources they need.]]

434 | "orgMember": { "resourceLinks": [{ "key": "getting-started" | "linkDisplayName": "Getting started guide" 436 | "url": "https://data.world/docs.data.world/en/59261-60585-0--Overview-of-data-world.html" 437 | "newTab": true 438 | "svgIcon": "meta" 439 | "verticallyAlignIcon": true } 440 | { "key": "documentation" 441 | "linkDisplayName": "Documentation" 442 | "url": "https://data.world/docs.data.world/" 443 | "newTab": true 444 | "svgIcon": "meta" 445 | "verticallyAlignIcon": true } 446 | { "key": "integrations" 447 | "linkDisplayName": "Integrations" 448 | "url": "https://data.world/integrations" 449 | "newTab": true 450 | "svgIcon": "integration" 451 | "verticallyAlignIcon": true 452 | "showIfAnyDismissed": ["onboard-checklist" 453 | "onboard-checklist-admin"] } 454 | { "key": "new-features" 455 | "linkDisplayName": "New features" 456 | "url": "https://data.world/whatnew.data.world/" 457 | "newTab": true 458 | "svgIcon": "magic-wand" 459 | "verticallyAlignIcon": true 460 | "showIfAnyDismissed": ["onboard-checklist" 461 | "onboard-checklist-admin"] } 462 | { "key": "enterprise" 463 | "linkDisplayName": "data.world for the enterprise" 464 | "url": "https://data.world/schedule-a-demo?src=inapp_data.world-for-the-enterprise" 465 | "newTab": true 466 | "svgIcon": "organization" 467 | "verticallyAlignIcon": true }] 468 | "onboardChecklist": { "key": "site" 469 | "items": [{ "id": "joinOrg" | "analyticsId": "onboard_checklist.join-org" | "href": "https://data.world/settings/organizations" | "completeIfMemberOfOrg": true | "formattedMessage": { "id": "onboard_checklist.join-org" | "description": "Text for onboard checklist join org link" | "defaultMessage": "Join organization" } } | { "id": "takeTour" 477 | "analyticsId": "onboard_checklist.take_tour" 478 | "showOnboardModalOnClick": true 479 | "completeIfDismissed": "onboard-tour-complete" 480 | "formattedMessage": { "id": "onboard_checklist.take_tour" 481 | "description": "Text for onboard checklist take tour link" 482 | "defaultMessage": "Take the tour" } } 483 | { "id": "customizeProfile" 484 | "analyticsId": "onboard_checklist.customize_profile" 485 | "href": "https://data.world/settings" 486 | "dismissOnClick": "customize-profile-link" 487 | "completeIfDismissed": "customize-profile-link" 488 | "formattedMessage": { "id": "onboard_checklist.customize_profile" 489 | "description": "Text for onboard checklist customize profile link" 490 | "defaultMessage": "Upload your photo" } } 491 | { "id": "createProject" 492 | "analyticsId": "onboard_checklist.create-project" 493 | "href": "https://data.world/create-a-project" 494 | "dismissOnClick": "create-project-link" 495 | "completeIfDismissed": "create-project-link" 496 | "formattedMessage": { "id": "onboard_checklist.create-project" 497 | "description": "Text for onboard checklist create project link" 498 | "defaultMessage": "Create your first project" } } 499 | { "id": "exploreIntegrations" 500 | "analyticsId": "onboard_checklist.explore-integrations" 501 | "href": "https://data.world/integrations" 502 | "target": "_blank" 503 | "dismissOnClick": "explore-integrations-link" 504 | "completeIfDismissed": "explore-integrations-link" 505 | "formattedMessage": { "id": "onboard_checklist.explore-integrations" 506 | "description": "Text for onboard checklist explore integrations link" 507 | "defaultMessage": "Explore integrations" } } } 508 | "onboardTour": [{ "show": true 509 | "header": "Welcome to data.world!" 510 | "subhead": "The place for modern data teamwork." 511 | "leftCol": "Organize data work so you and your team find what you

need faster."512|"middleCol":"Explore and create a shared understanding of your data."513|"rightCol":"Share insights and invite more people to participate in data teamwork."}514|{"show":true515|"imageHeaderKey":"search"516|"imageKey":"collaborate"517|"imageAlt":"Start with Data"518|"header":"Start with fantastic data"519|"subhead":"Browse thousands of free520| open data resources; bookmark your favorites for later.\n\nGather your own data from almost anywhere into a single searchable library.\n\nMake it open521| share with your team522| or keep it private."}523|{"show":true524|"imageHeaderKey":"bulb"525|"imageKey":"dataProject"526|"imageAlt":"Data Projects"527|"header":"Organize your data projects"528|"subhead":"Create projects that hold everything you need to answer critical questions.\n\nAdd context to make it easier for others to discover529| understand530| and contribute to your projects.\n\nImmediately query data without leaving your projects."}531|{"show":true532|"imageHeaderKey":"code"533|"imageKey":"slideIntegrations"534|"imageAlt":"Integrations"535|"reverse":true536|"header":"Connect the data tools you love"537|"subhead":"Automatically keep data.world synced with your data sources.\n\nIntegrate your team's favorite tools538| from analysis to visualization."}539|{"show":true540|"imageHeaderKey":"discussions"541|"imageKey":"discussionsCollaboration"542|"imageAlt":"Contributors"543|"header":"Add more brainpower to your projects"544|"subhead":"Invite your team to securely and effectively work together.\n\nManage fine-grained permissions for each member.\n\nAsk questions545| add insights546| bookmark top contributions547| and more."}548|{"show":true549|"imageHeaderKey":"insight"550|"imageKey":"knowYourData"551|"imageAlt":"know your data"552|"header":"Share your results"553|"subhead":"Share meaningful insights and the data behind them.\n\nDistribute clean554| documented data for others to use in their own work.\n\nHelp the right people discover the resources they need."}}555|"other":{"resourceLinks":[{"key":"getting-started|"linkDisplayName":"Getting started guide"557|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"558|"newTab":true559|"svgIcon":"meta"560|"verticallyAlignIcon":true}561|{"key":"documentation"562|"linkDisplayName":"Documentation"563|"url":"https:\u002F\u002Fdocs.data.world\u002F"564|"newTab":true565|"svgIcon":"meta"566|"verticallyAlignIcon":true}567|{"key":"integrations"568|"linkDisplayName":"Integrations"569|"url":"\u002Fintegrations"570|"newTab":true571|"svgIcon":"integration"572|"verticallyAlignIcon":true573|"showIfAnyDismissed":["onboard-checklist"574|"onboard-checklist-admin"]}575|{"key":"new-features"576|"linkDisplayName":"New features"577|"url":"https:\u002F\u002Fwhatsnew.data.world"578|"newTab":true579|"svgIcon":"magic-wand"580|"verticallyAlignIcon":true581|"showIfAnyDismissed":["onboard-checklist"582|"onboard-checklist-admin"]}583|{"key":"enterprise"584|"linkDisplayName":"data.world for the enterprise"585|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-demo\u002F?src=inapp_data.world-for-the-enterprise"586|"newTab":true587|"svgIcon":"organization"588|"verticallyAlignIcon":true}589|"enableImageUpload":true|"onboardChecklist":{"key":"site"591|"items":[{"id":"createAccount"592|"analyticsId":"onboard_checklist.create_account"593|"complete":true594|"formattedMessage":{"id":"onboard_checklist.create_account"595|"description":"Text for onboard

checklist create account link"596|"defaultMessage":"Create account"}}597|{"id":"takeTour"598|"analyticsId":"onboard_checklist.take_tour"599|"showOnboardModalOnClick":true600|"completeIfDismissed":"onboard-tour-complete"601|"formattedMessage":{"id":"onboard_checklist.take_tour"602|"description":"Text for onboard checklist take tour link"603|"defaultMessage":"Take the tour"}}604|{"id":"customizeProfile"605|"analyticsId":"onboard_checklist.customize_profile"606|"href":"\u002Fsettings"607|"dismissOnClick":"customize-profile-link"608|"completeIfDismissed":"customize-profile-link"609|"formattedMessage":{"id":"onboard_checklist.customize_profile"610|"description":"Text for onboard checklist customize profile link"611|"defaultMessage":"Upload your photo"}}612|{"id":"createProject"613|"analyticsId":"onboard_checklist.create-project"614|"href":"\u002Fcreate-a-project"615|"dismissOnClick":"create-project-link"616|"completeIfDismissed":"create-project-link"617|"formattedMessage":{"id":"onboard_checklist.create-project"618|"description":"Text for onboard checklist create project link"619|"defaultMessage":"Create your first project"}}620|{"id":"exploreIntegrations"621|"analyticsId":"onboard_checklist.explore-integrations"622|"href":"\u002Fintegrations"623|"target":"_blank"624|"dismissOnClick":"explore-integrations-link"625|"completeIfDismissed":"explore-integrations-link"626|"formattedMessage":{"id":"onboard_checklist.explore-integrations"627|"description":"Text for onboard checklist explore integrations link"628|"defaultMessage":"Explore integrations"}}}}629|"onboardTour":[{"show":true630|"header":"Welcome to data.world!"631|"subhead":"The place for modern data teamwork."632|"leftCol":"Organize data work so you and your team find what you need faster."633|"middleCol":"Explore and create a shared understanding of your data."634|"rightCol":"Share insights and invite more people to participate in data teamwork."}635|{"show":true636|"imageHeaderKey":"search"637|"imageKey":"collaborate"638|"imageAlt":"Start with Data"639|"header":"Start with fantastic data"640|"subhead":"Browse thousands of free641| open data resources; bookmark your favorites for later.\n\nGather your own data from almost anywhere into a single searchable library.\n\nMake it open642| share with your team643| or keep it private."}644|{"show":true645|"imageHeaderKey":"bulb"646|"imageKey":"dataProject"647|"imageAlt":"Data Projects"648|"header":"Organize your data projects"649|"subhead":"Create projects that hold everything you need to answer critical questions.\n\nAdd context to make it easier for others to discover650| understand651| and contribute to your projects.\n\nImmediately query data without leaving your projects."}652|{"show":true653|"imageHeaderKey":"code"654|"imageKey":"slideIntegrations"655|"imageAlt":"Integrations"656|"reverse":true657|"header":"Connect the data tools you love"658|"subhead":"Automatically keep data.world synced with your data sources.\n\nIntegrate your team's favorite tools659| from analysis to visualization."}660|{"show":true661|"imageHeaderKey":"discussions"662|"imageKey":"discussionsCollaboration"663|"imageAlt":"Contributors"664|"header":"Add more brainpower to your projects"665|"subhead":"Invite your team to securely and effectively work together.\n\nManage fine-grained permissions for each member.\n\nAsk questions666| add insights667| bookmark top contributions668| and more."}669|{"show":true670|"imageHeaderKey":"insi

```

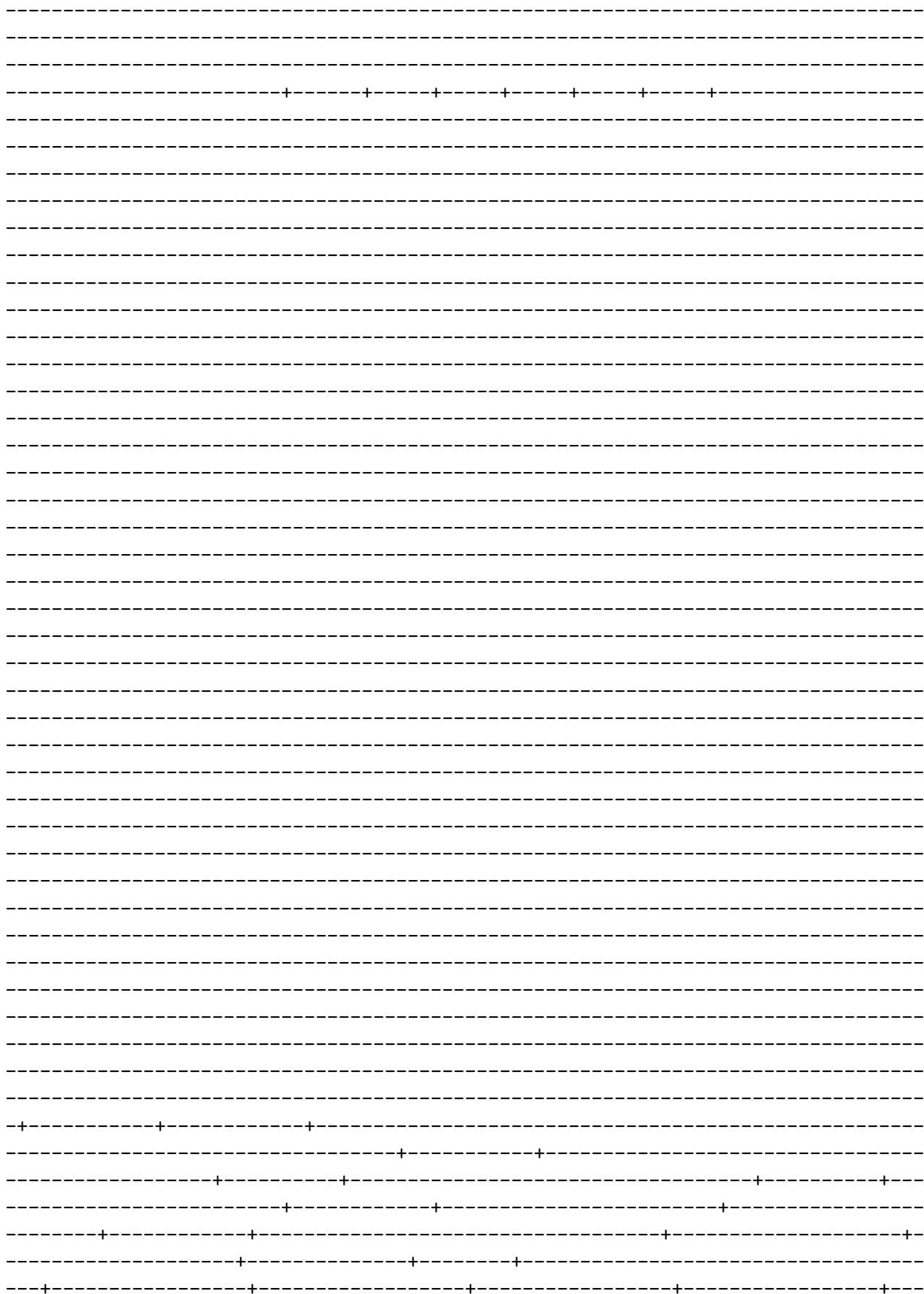
ght"671|"imageKey":"knowYourData"672|"imageAlt":"know your
data"673|"header":"Share your results"674|"subhead":"Share meaningful insights
and the data behind&nbsp;them.\n\nDistribute clean675| documented data for
others to use in their own&nbsp;work.\n\nHelp the right people discover the
resources they&nbsp;need."}}676|"newResourceItems":[{"type":"dataset|"exampleL
ink":"https:\u002F\u002Fdata.world\u002Fninja\u002Fanw-obstacle-history"}|{"type
":"project|"exampleLink":"https:\u002F\u002Fdata.world\u002Fmarkmarkoh\u002Fexp
loring-thor"}}|{"version_build":"20230921123426_7d6ba6c4|"version_check_interva
l":60000|"version_commit":"7d6ba6c4|"visibility":{"discoverableVisibilitySubHea
derText":"This \u003C%= resourceType %\u003E can be discovered by all data.world
users. You can change the \u003C%= resourceType %\u003E visibility or make it
private in the table below."|"openVisibilitySubHeaderText":"This \u003C%=
resourceType %\u003E can be discovered and viewed by anyone on data.world. You
can change the \u003C%= resourceType %\u003E visibility or make it private in
the table below."|"privateVisibilityShareButtonText":"Share with
community"|"privateVisibilitySubHeaderText":"This \u003C%= resourceType %\u003E
is only visible to you. It is neither discoverable nor viewable by the
data.world community. To open it up to others to view| change the visibility
settings below."|"shareModalBodyText":"Set the visibility level for all of
data.world below. The content of the \u003C%= resourceType %\u003E will only be
visible to those who have been granted explicit access to this \u003C%=
resourceType %\u003E."|"shareModalHeaderText":"Share with all
data.world"|"visibilityTableTitleCellText":"Anyone on
data.world"}|"context_id":"b04792da5b608cdd"}</script><script
type="application/json" id="init-profile">{"requestAccessForm":[{"label":"Full n
ame"|"required":true693|"readOnly":true694|"showOptionalFlag":false695|"lookupPr
operty":"displayName"696|"type":"input"}|{"label":"Contact
email"|"placeholder":"Provide an email in the case a dataset administrator needs
to reach out"699|"required":false|"readOnly":false701|"showOptionalFlag":true|"t
ype":"email"}|{"label":"Message"|"placeholder":"Share what you would like to do
with this dataset"705|"required":true706|"readOnly":false707|"showOptionalFlag":
false708|"type":"textarea"}]}</script><script type="application/json" id="init-
analytics">{"siteId":"00000"}</script><script src="https://js.stripe.com/v3/"
async id="stripe-js"></script> <script type="text/javascript" src="https://asse
ts.data.world/js/runtime.4c0777f218351a402063.js"></script><script
type="text/javascript" src="https://assets.data.world/js/vendors~app.73c77cf6388
8ff69bcb2.js"></script><script type="text/javascript" src="https://assets.data.w
orld/js/vendors~app~vendors.66b69a1400c3cdba7c1a.js"></script><script
type="text/javascript" src="https://assets.data.world/js/vendors~vendors.595dac1
9a1bcced9e705.js"></script><script type="text/javascript" src="https://assets.da
ta.world/js/vendors.6dfd2438038a5ca9e317.js"></script><script
type="text/javascript" src="https://assets.data.world/js/app.15687b4170eaf580bb0
2.js"></script></body></html>|

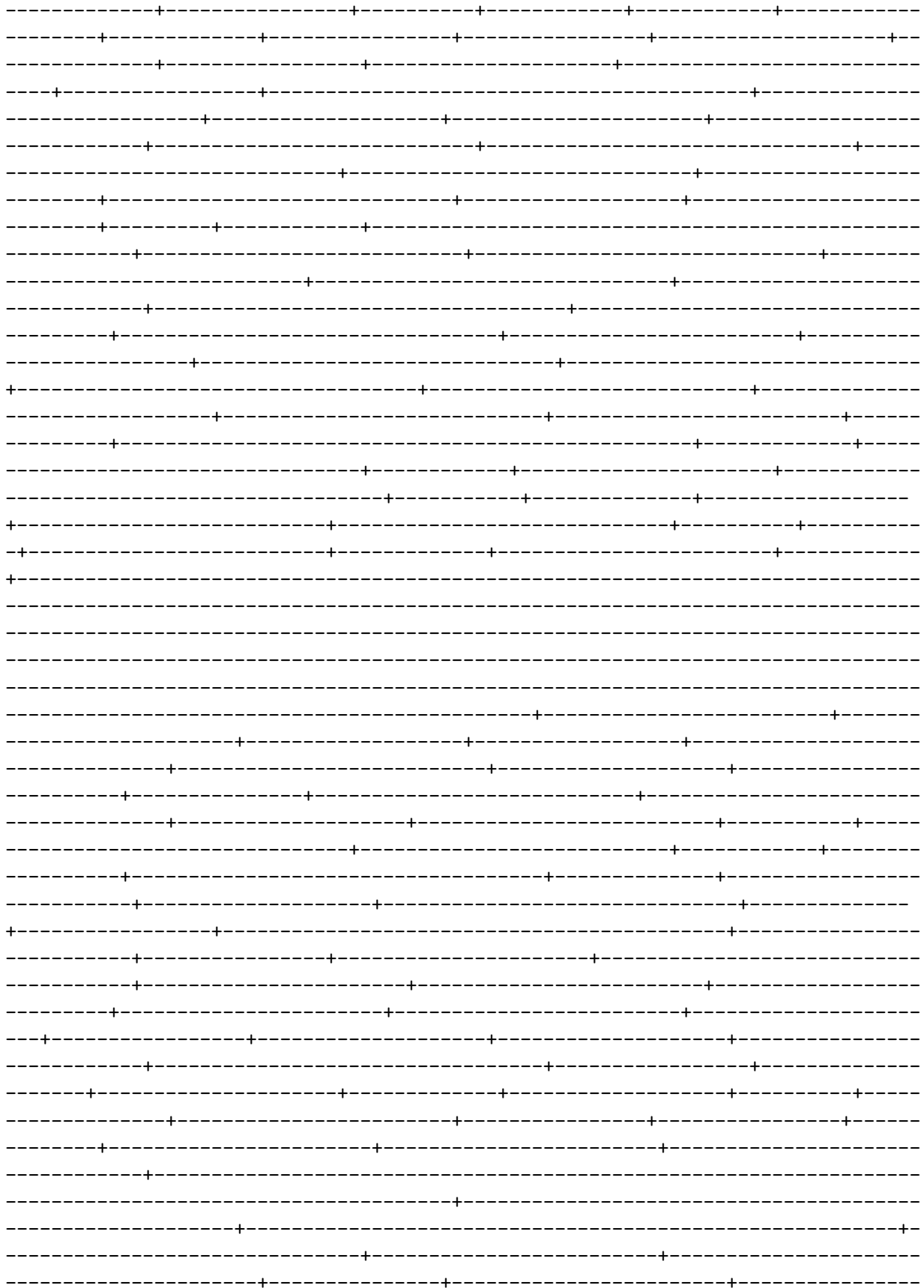
```

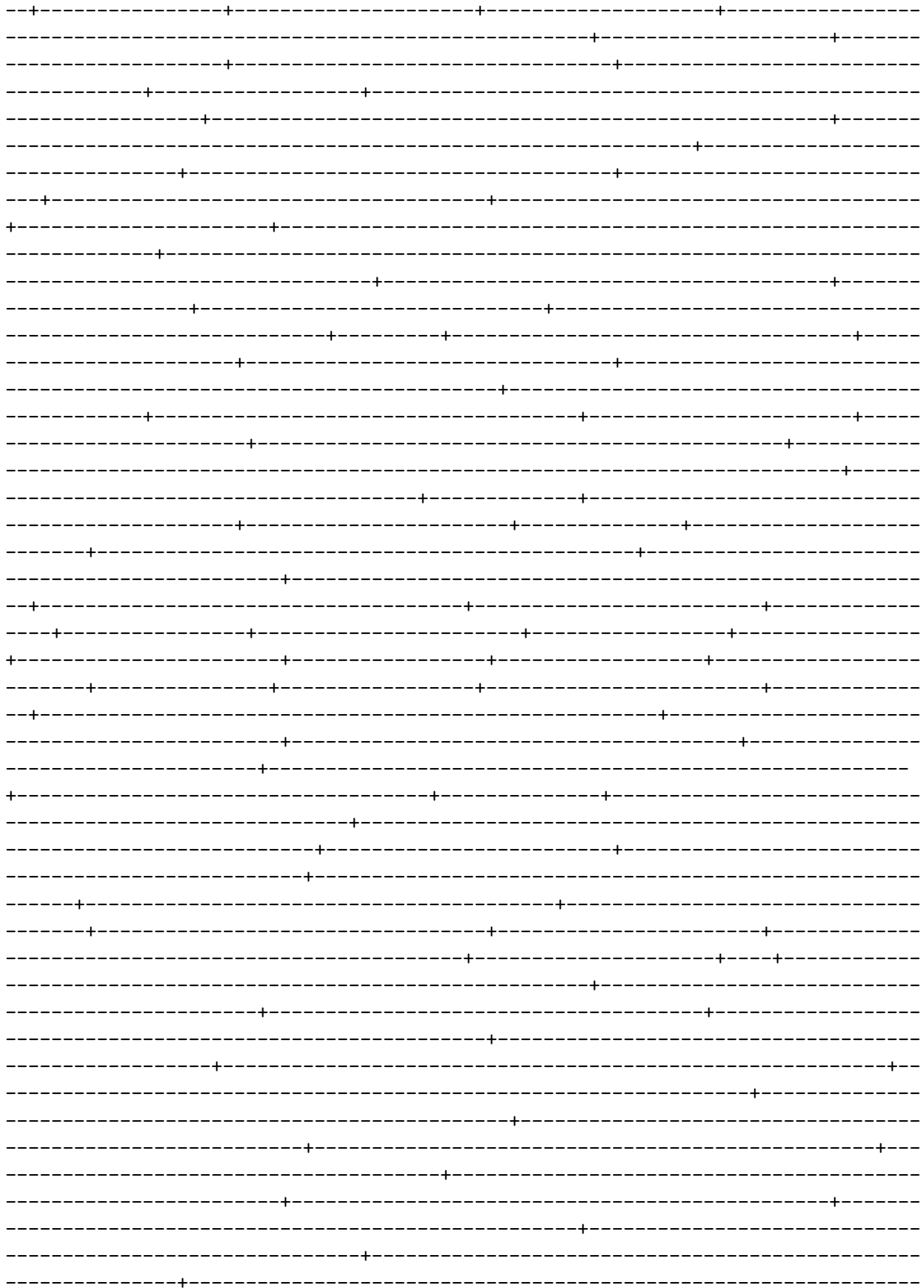
```

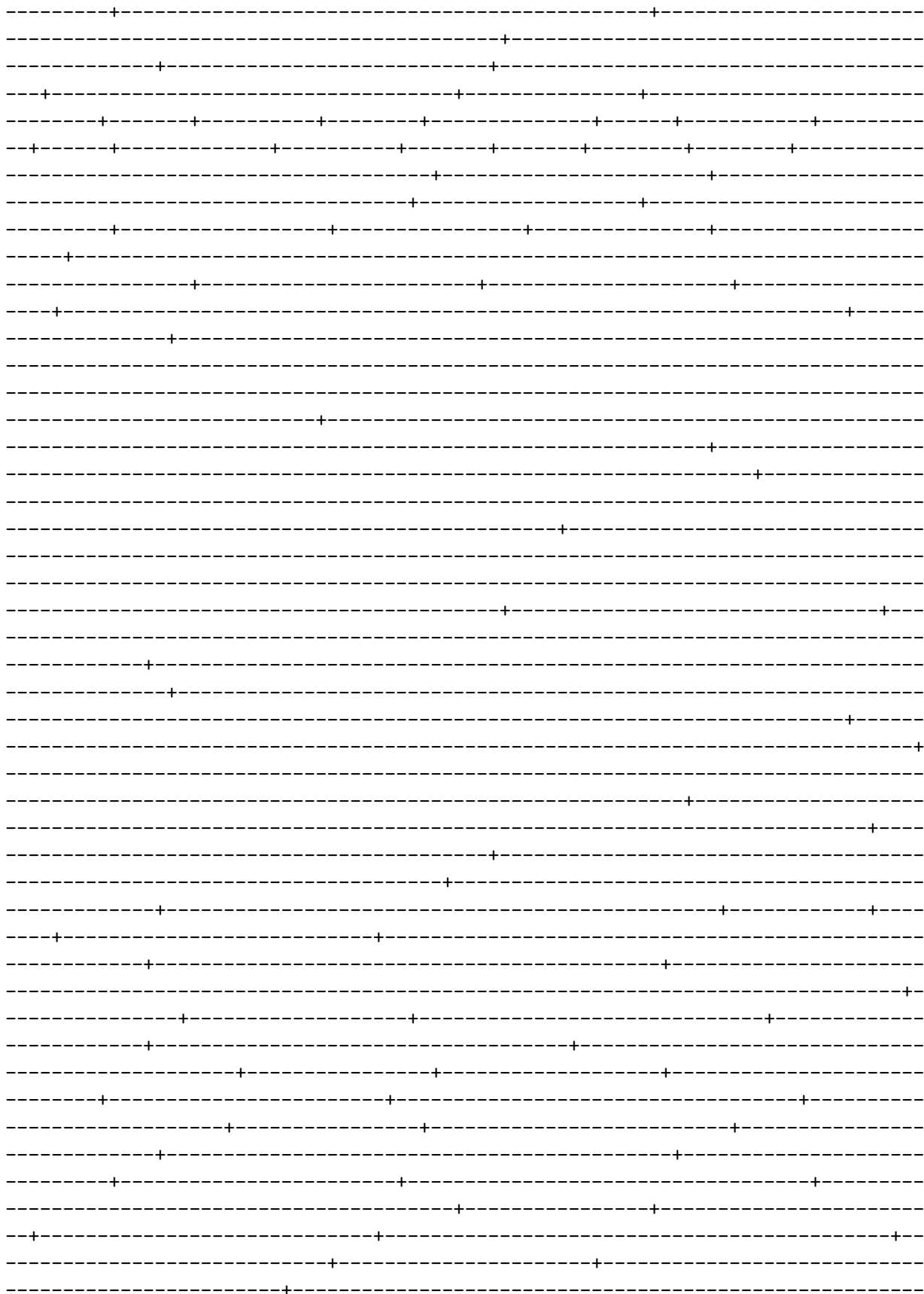
+-----+
+-----+
+-----+
+-----+
+-----+

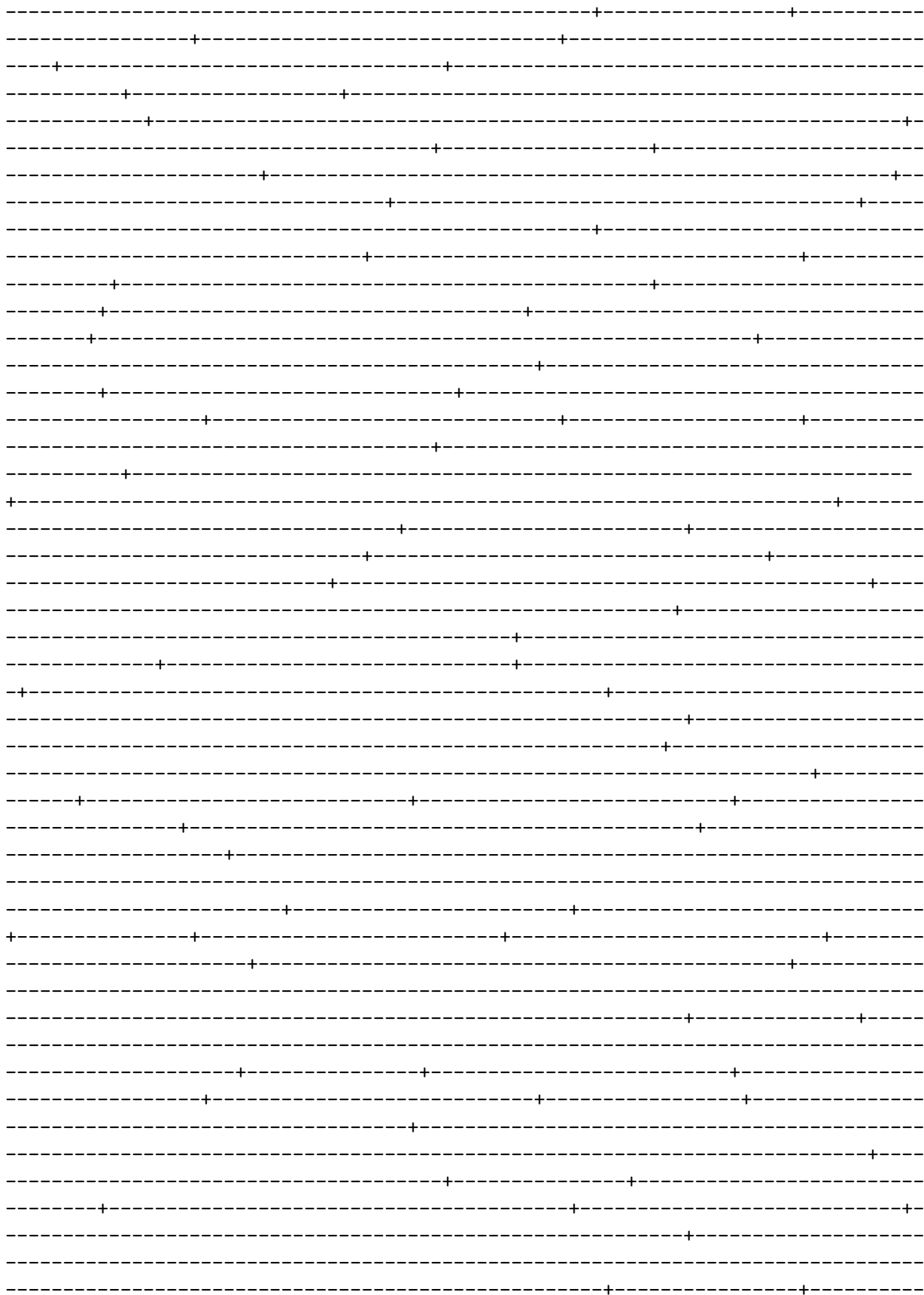
```

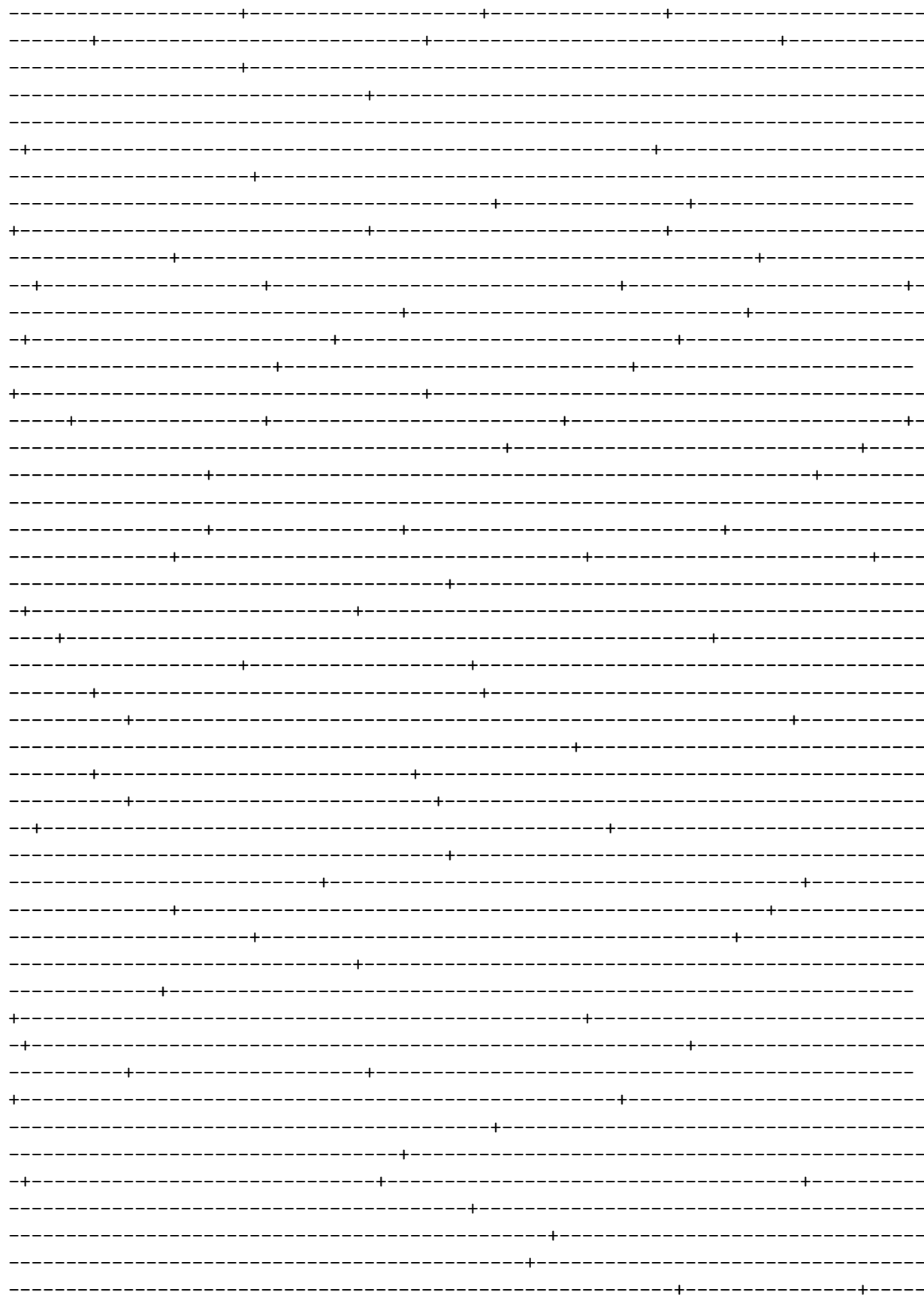


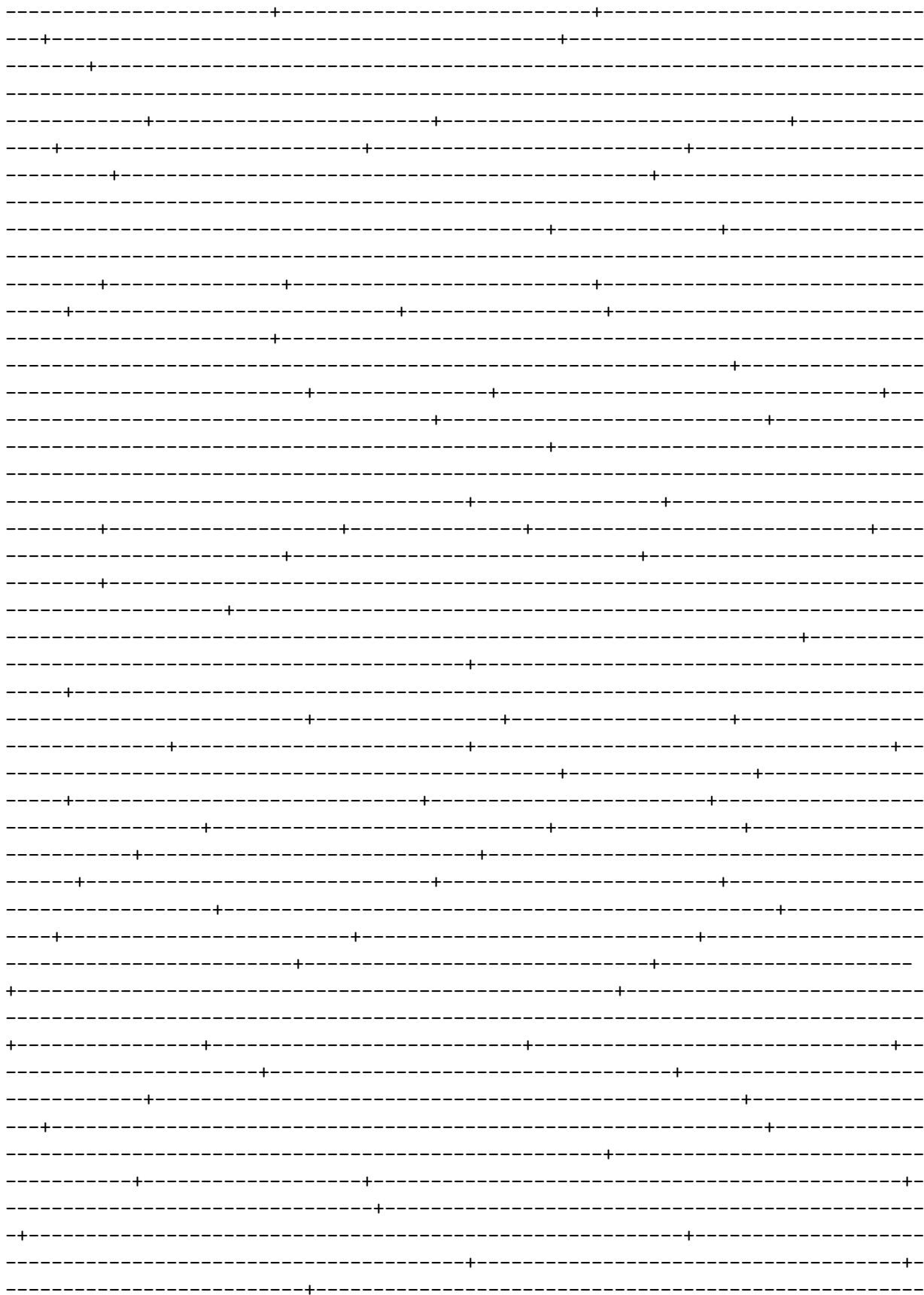


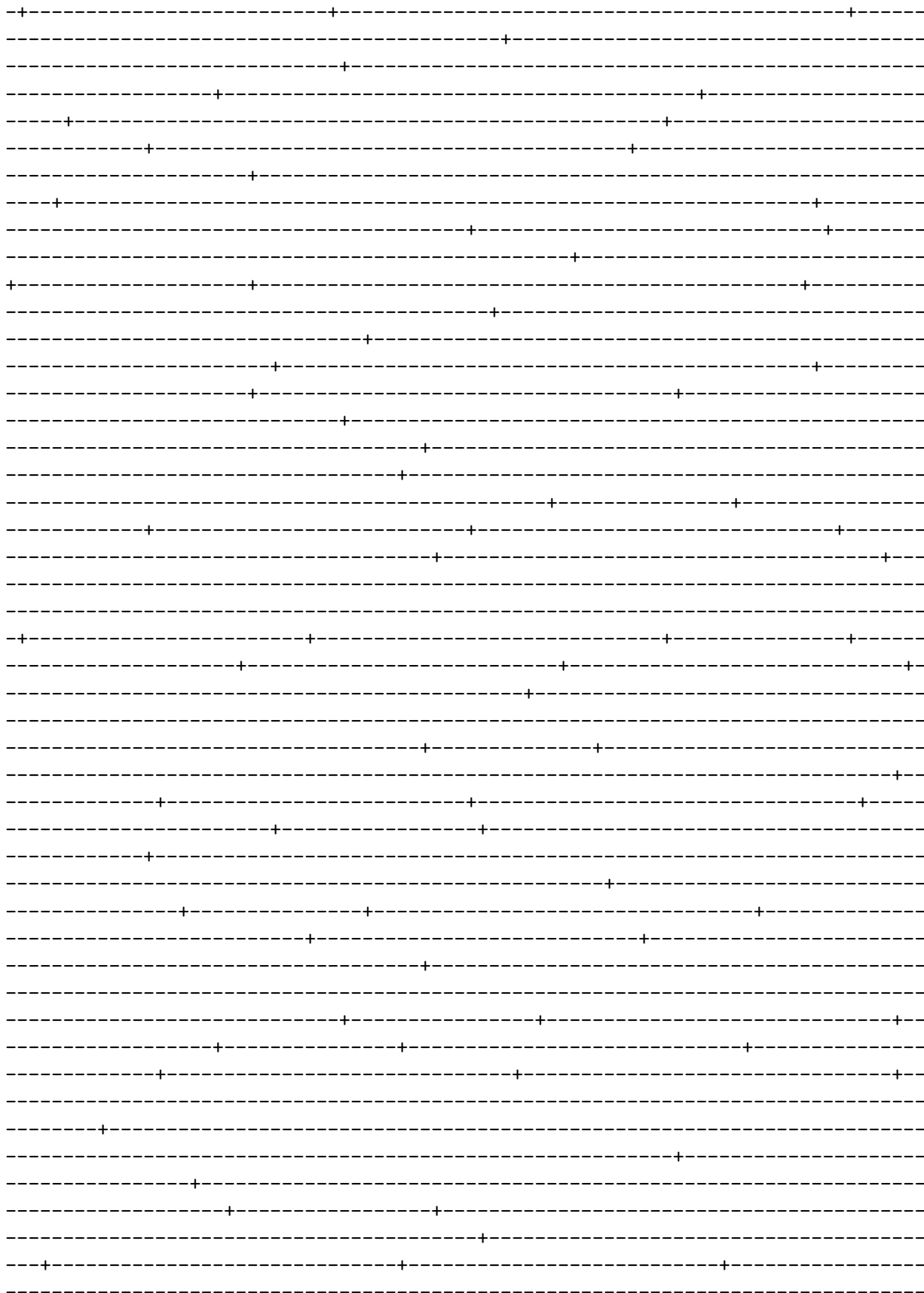


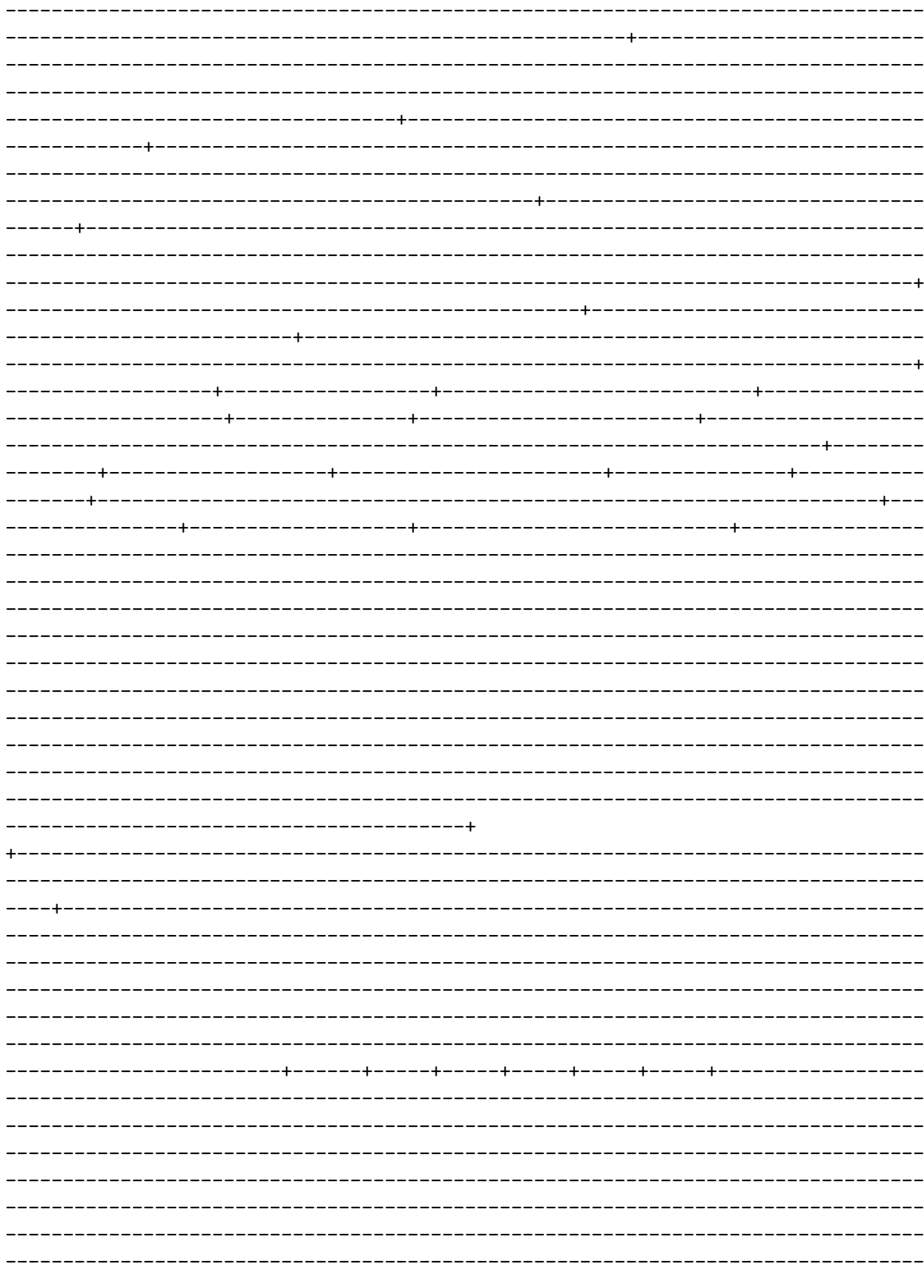


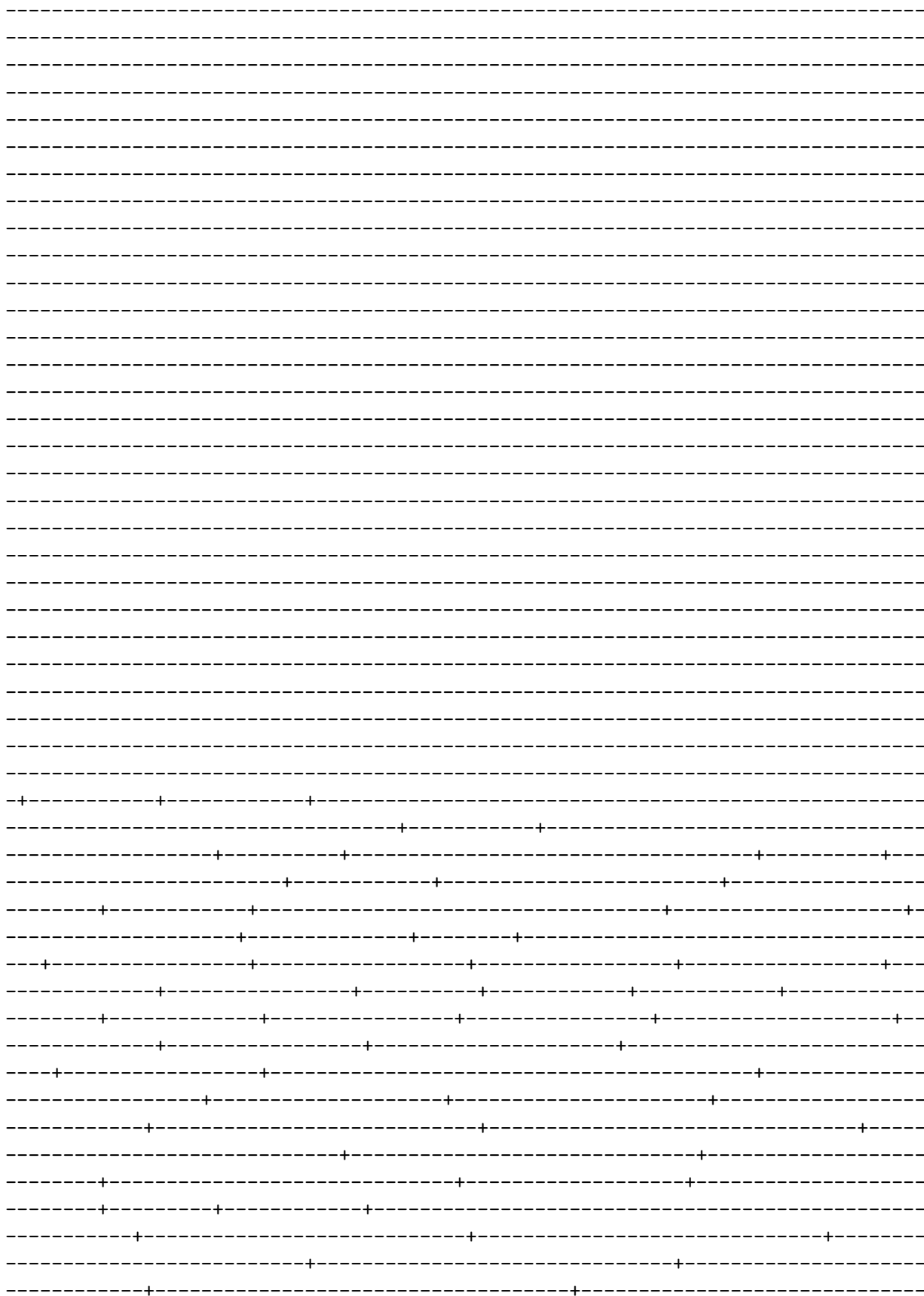


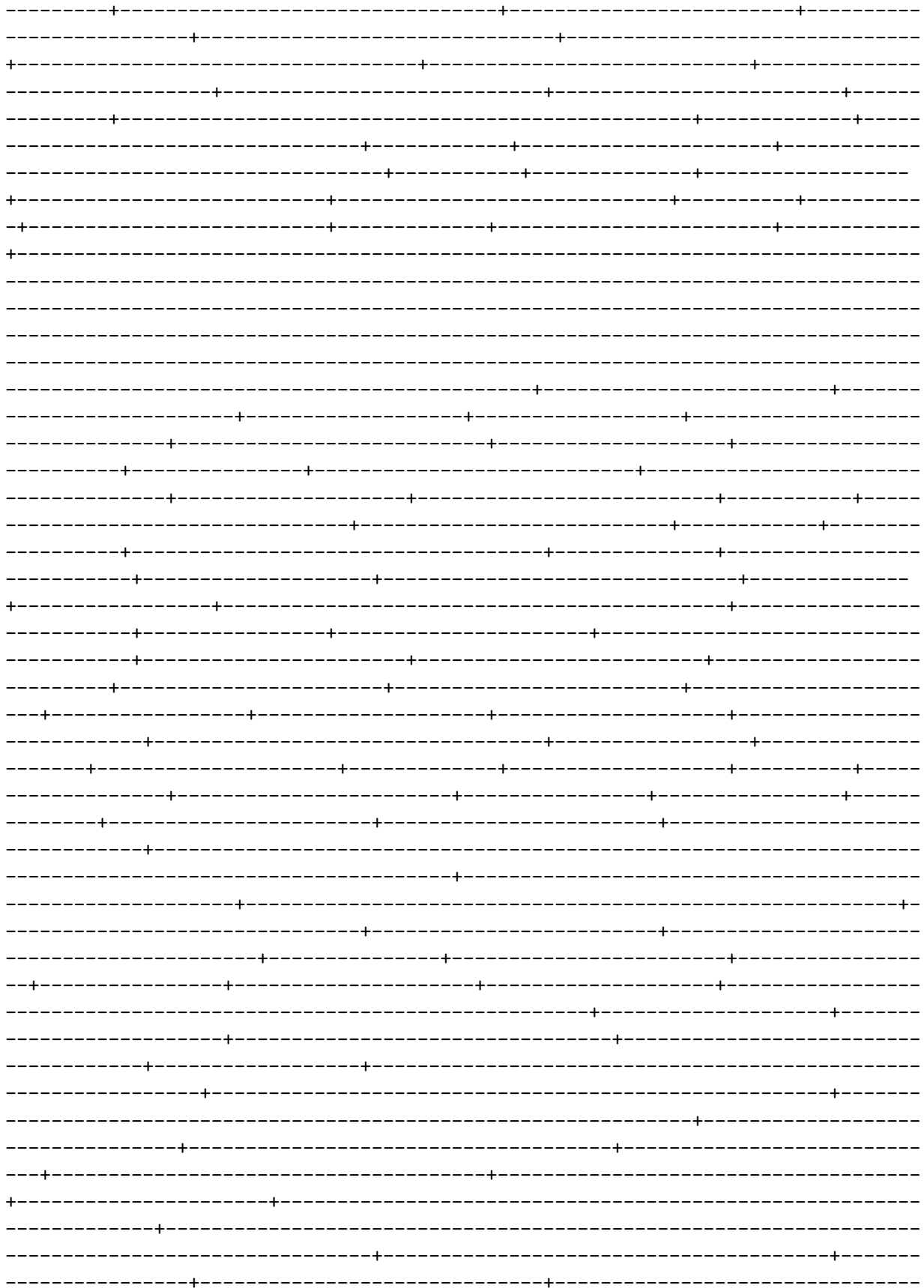


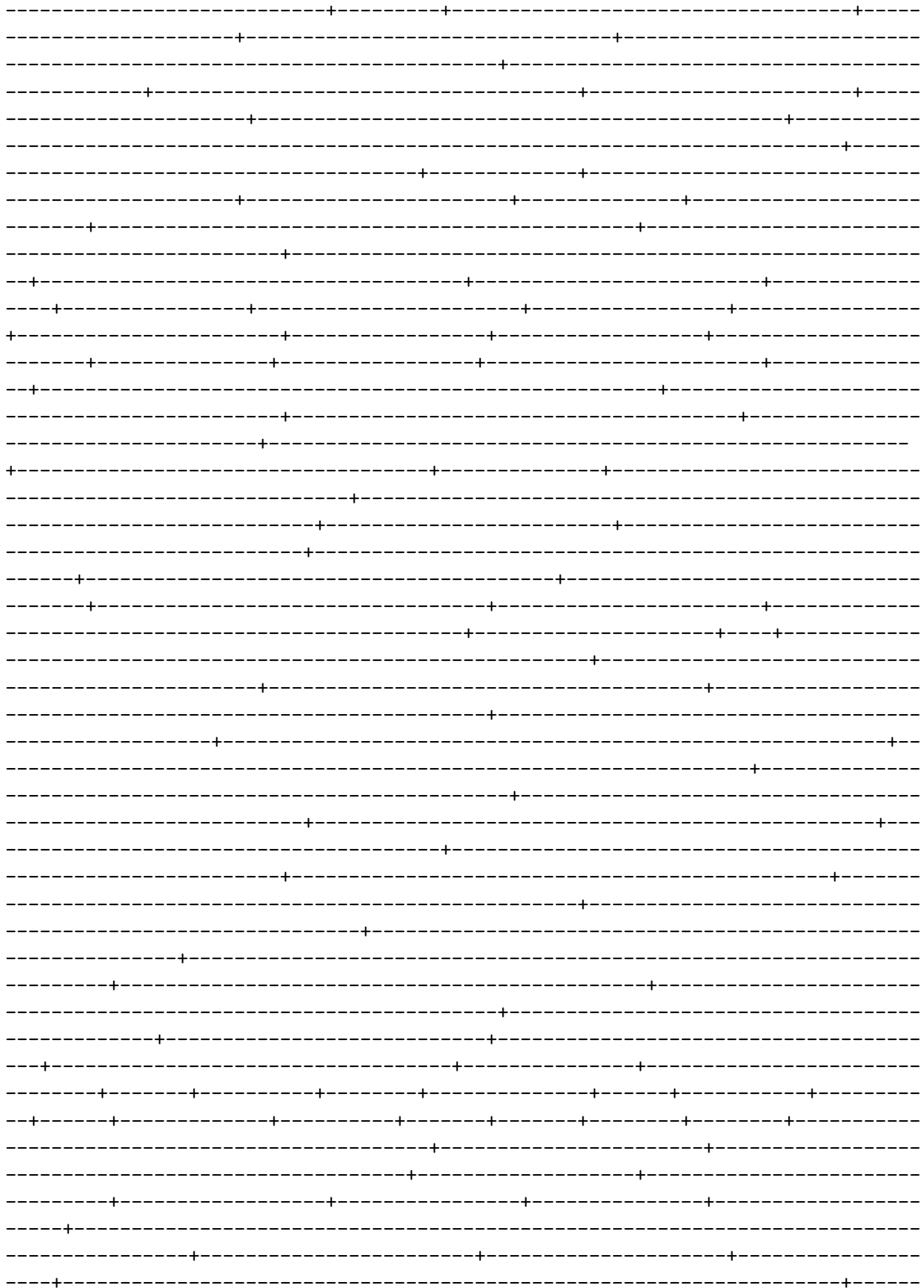


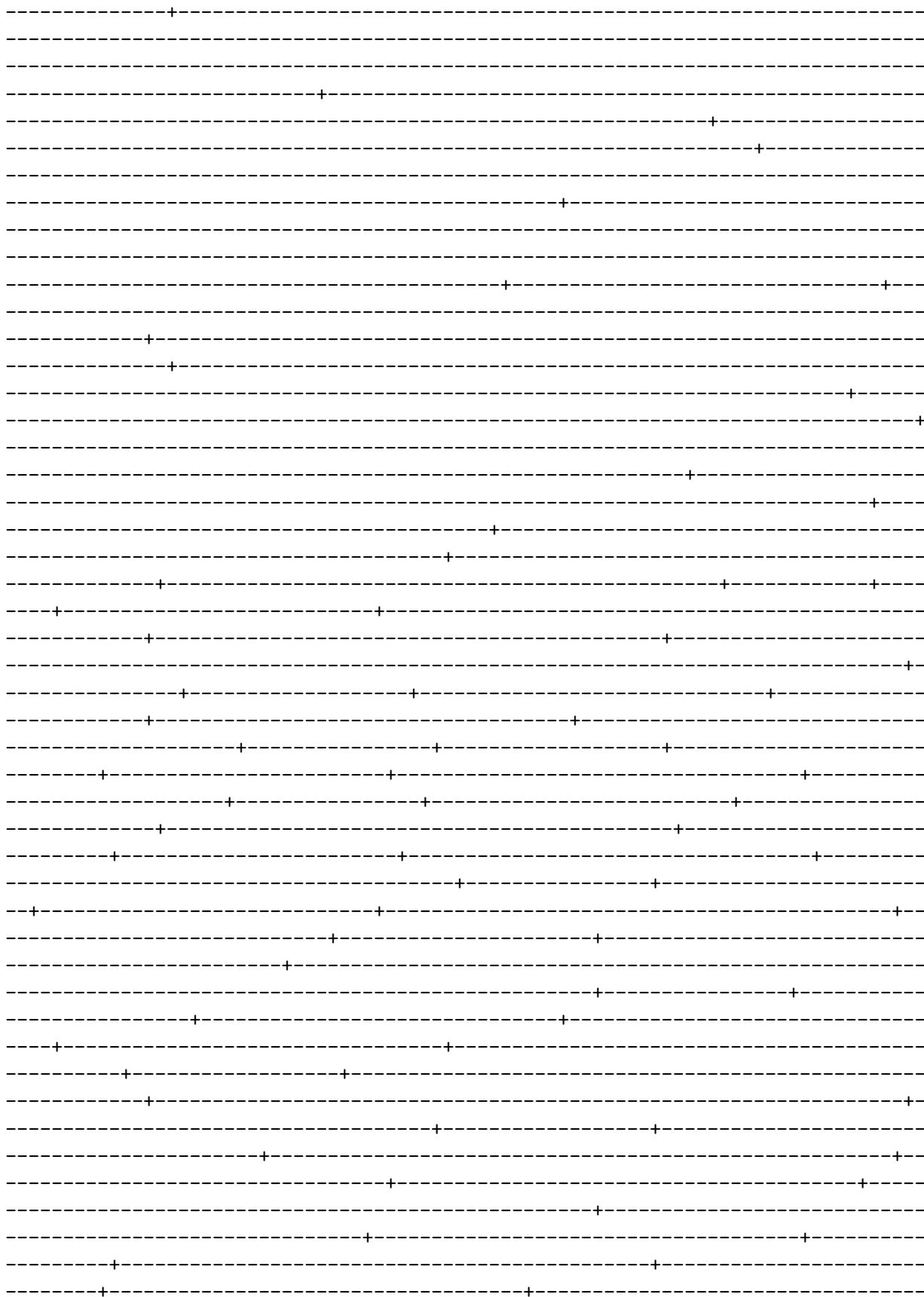


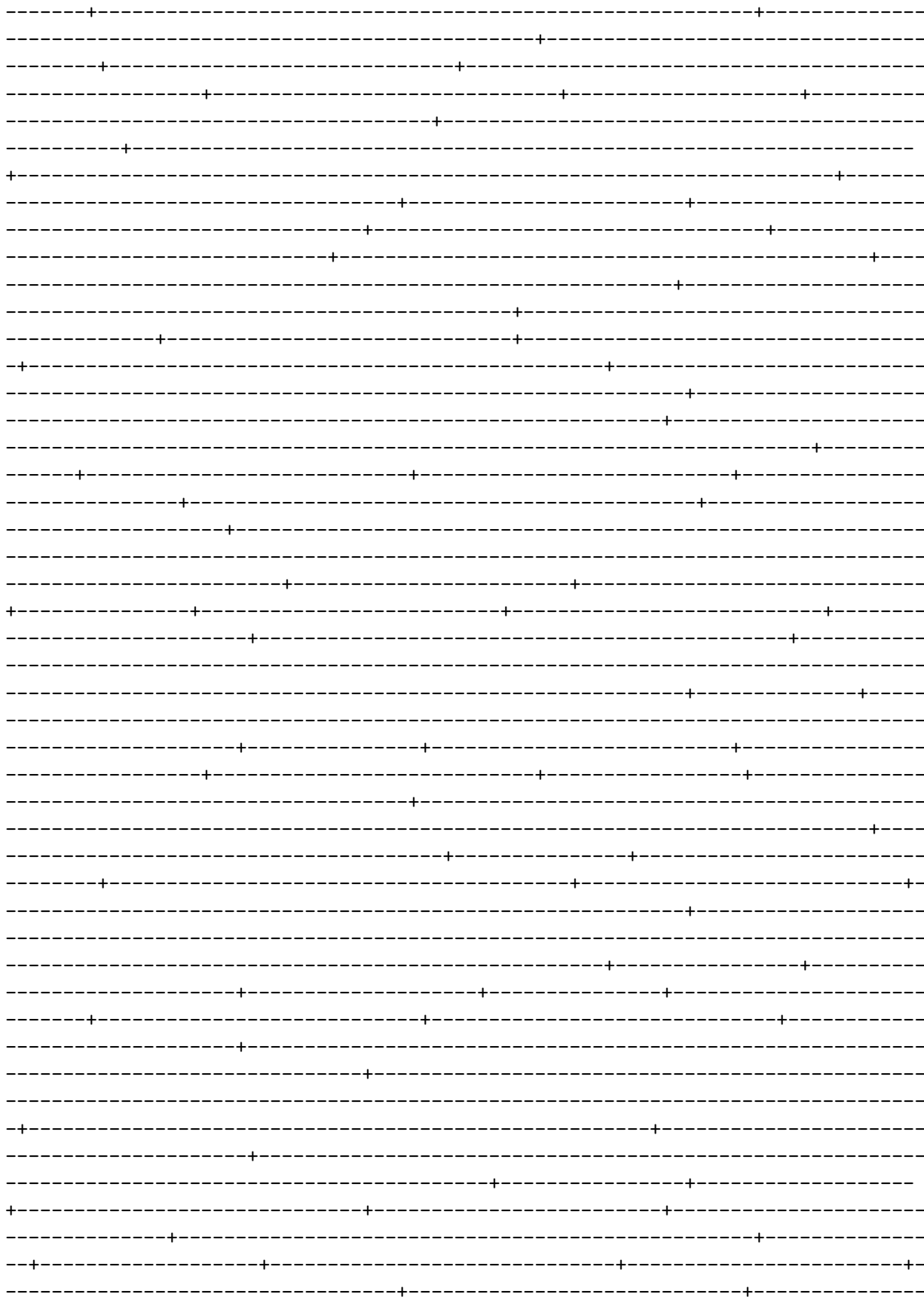


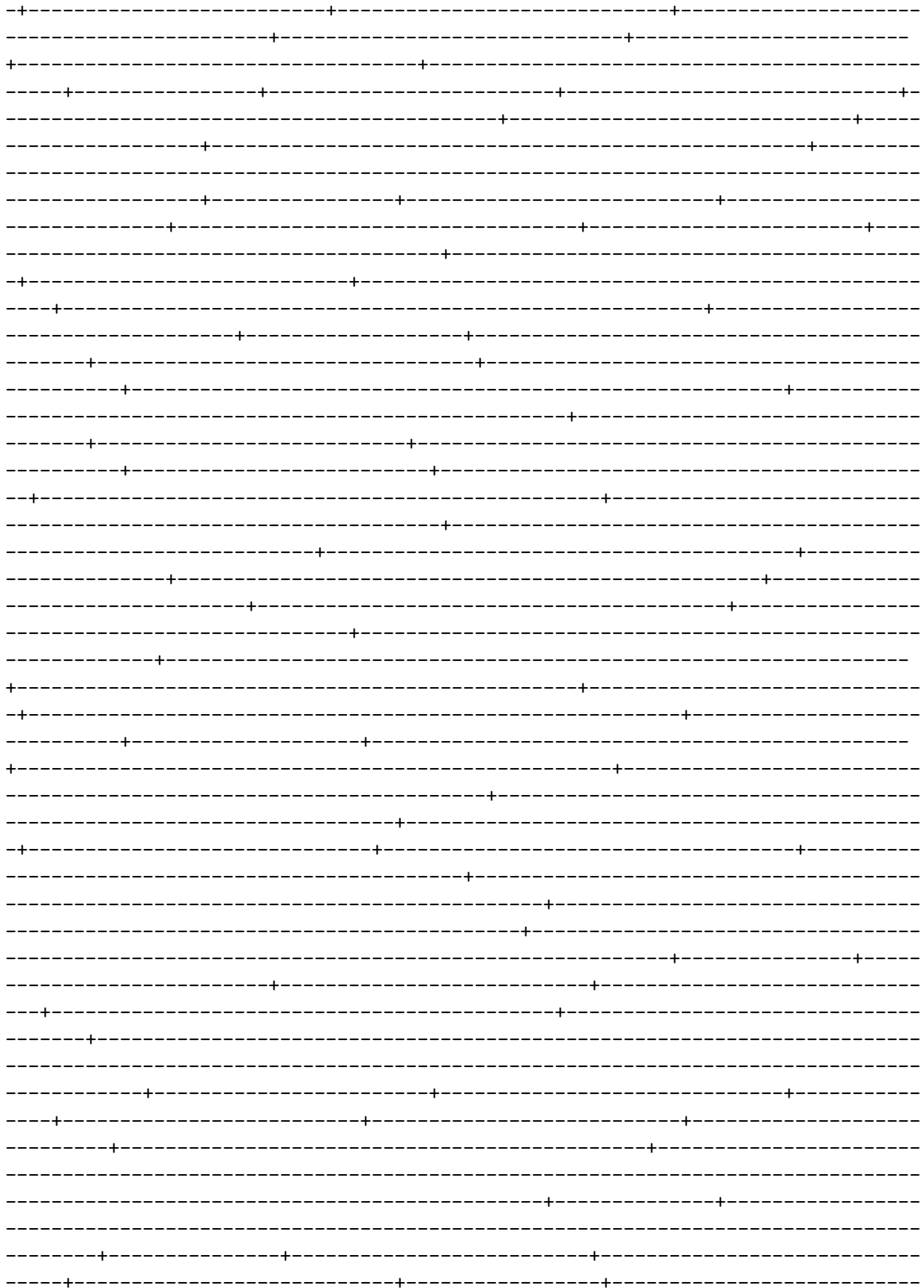


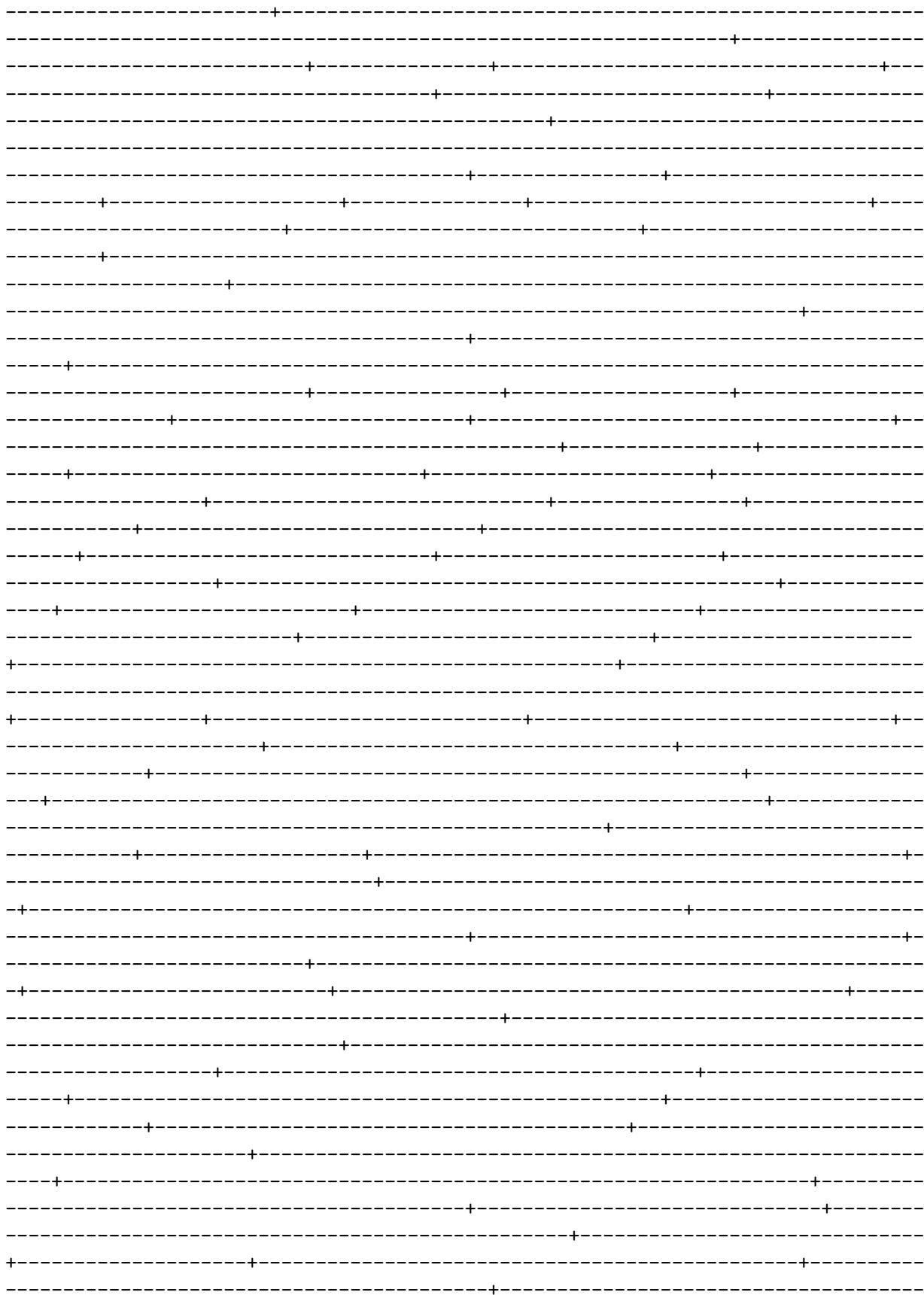


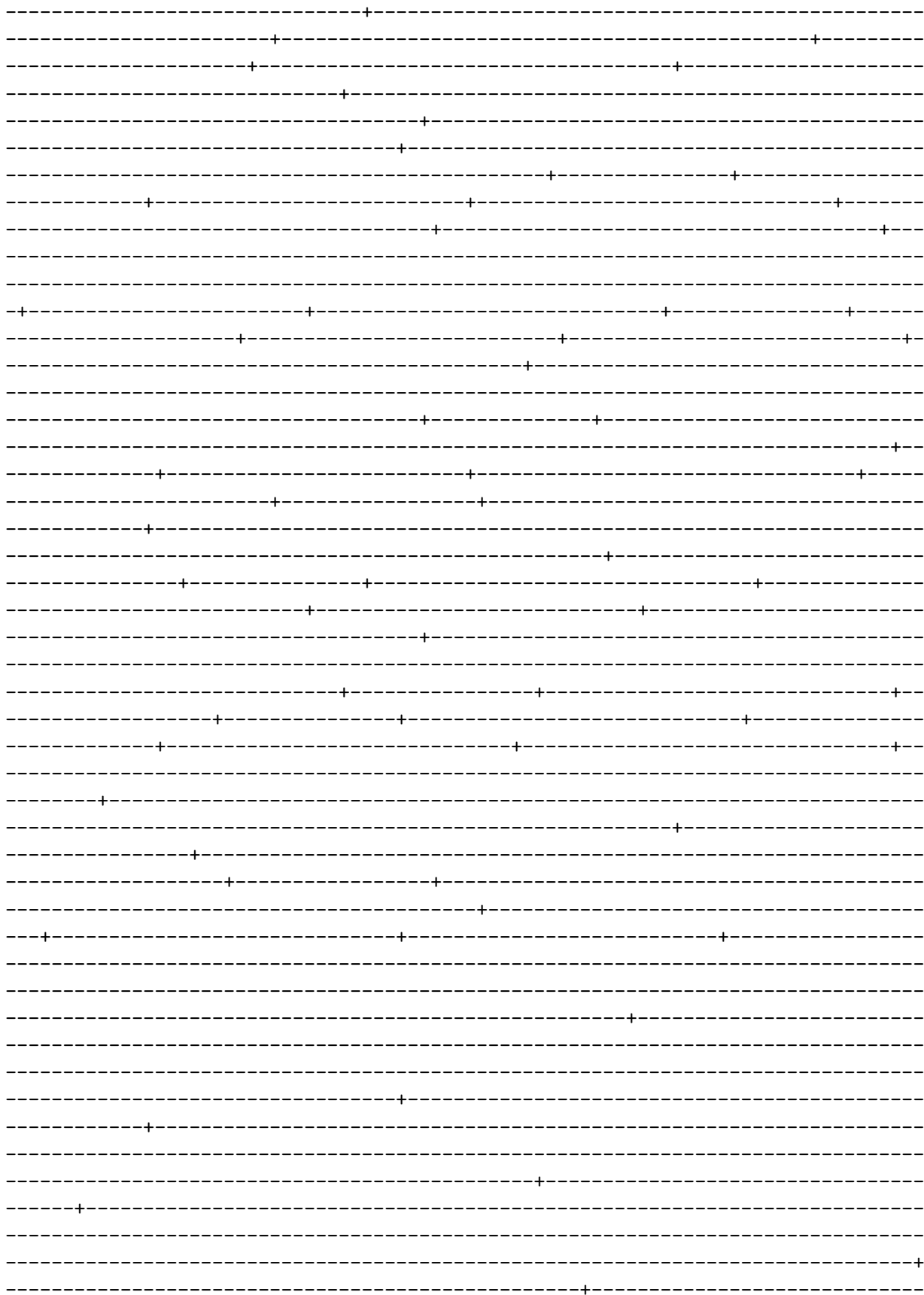


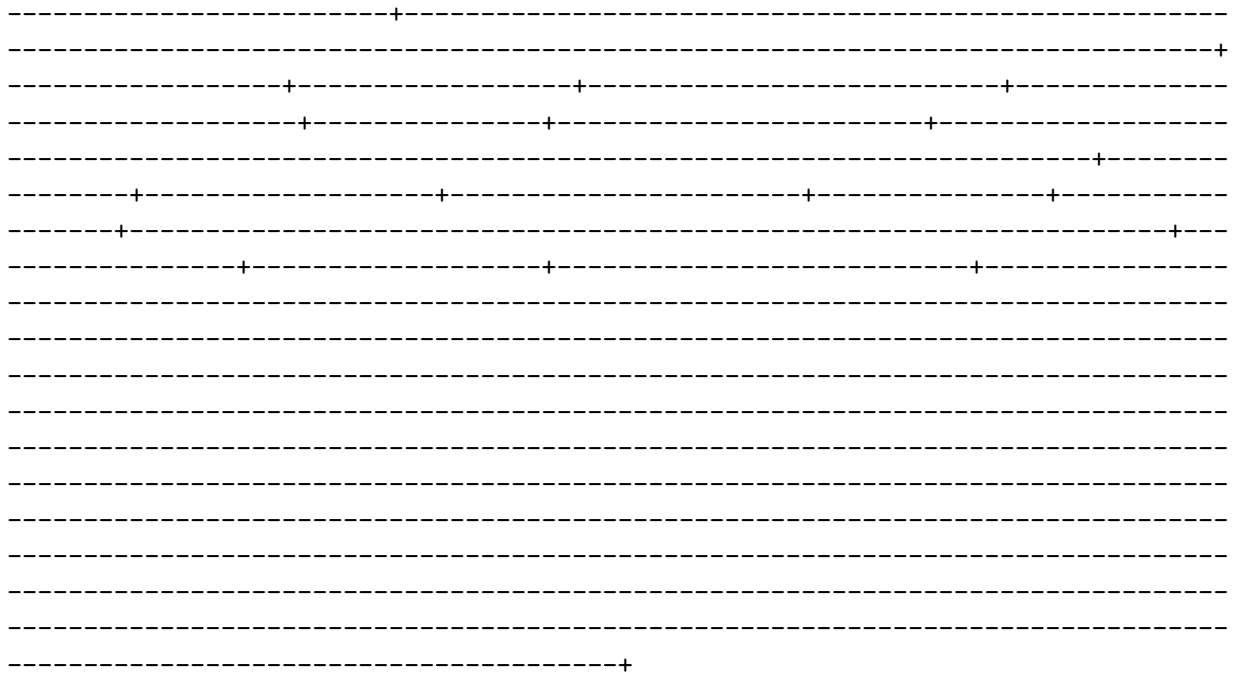












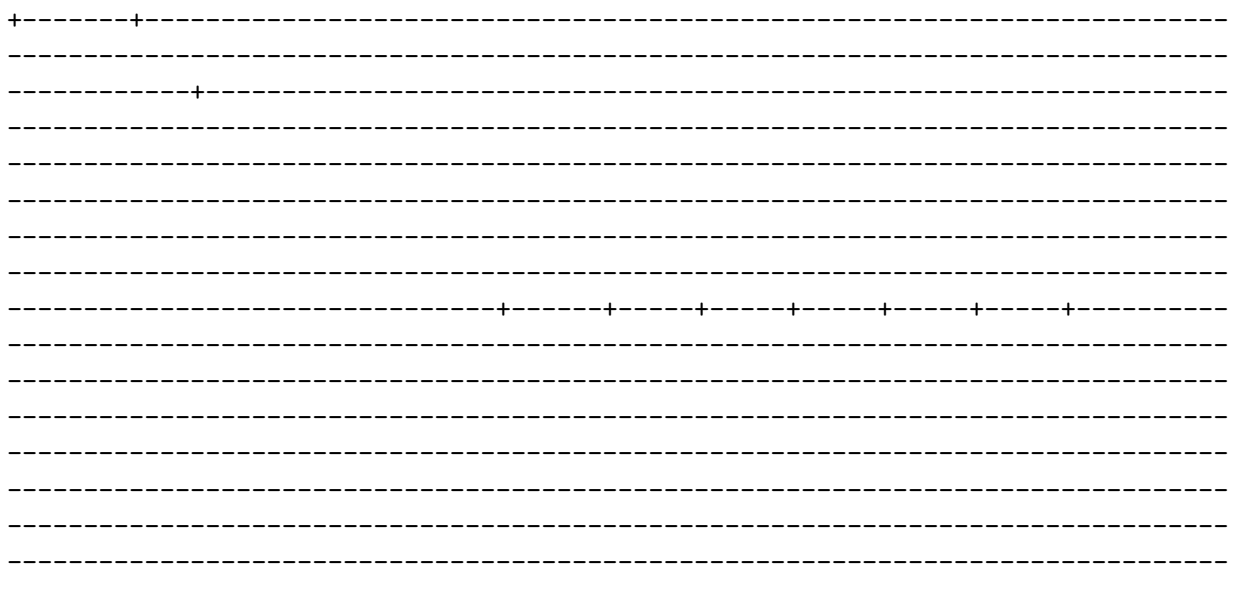
```
[34]: british2_df.count()
```

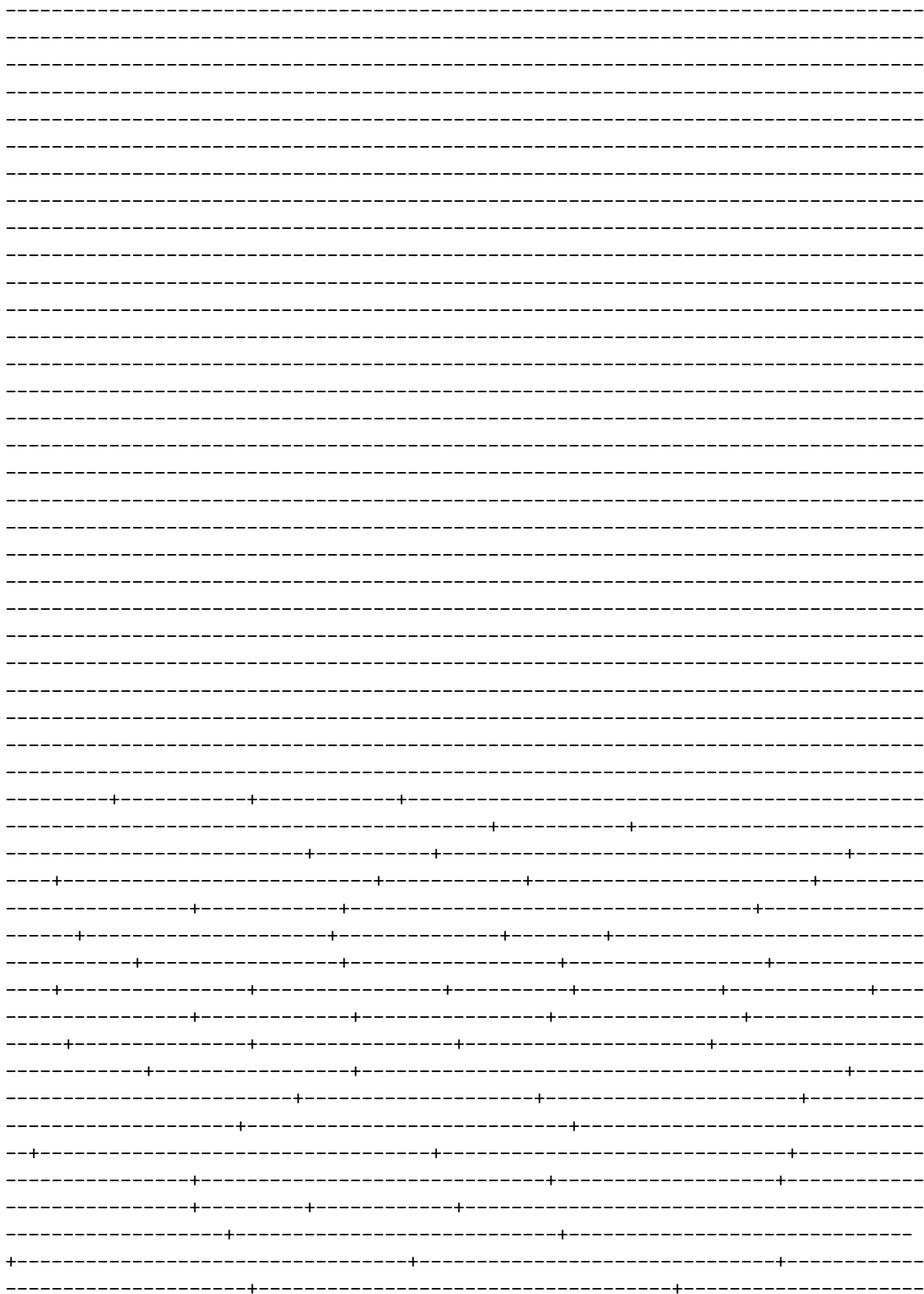
```
[34]: 0
```

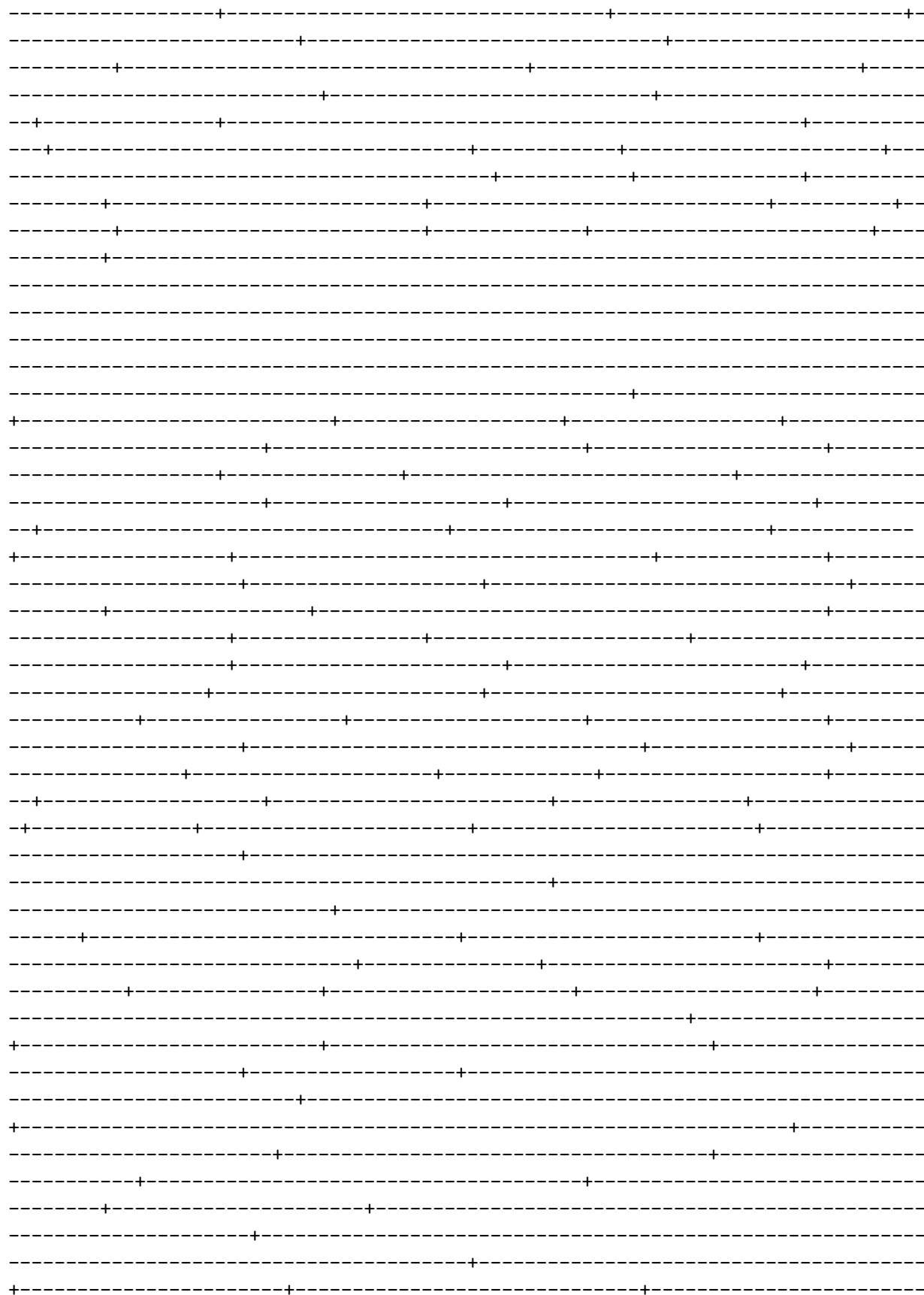
```
[35]: british2_df.describe().show()
```

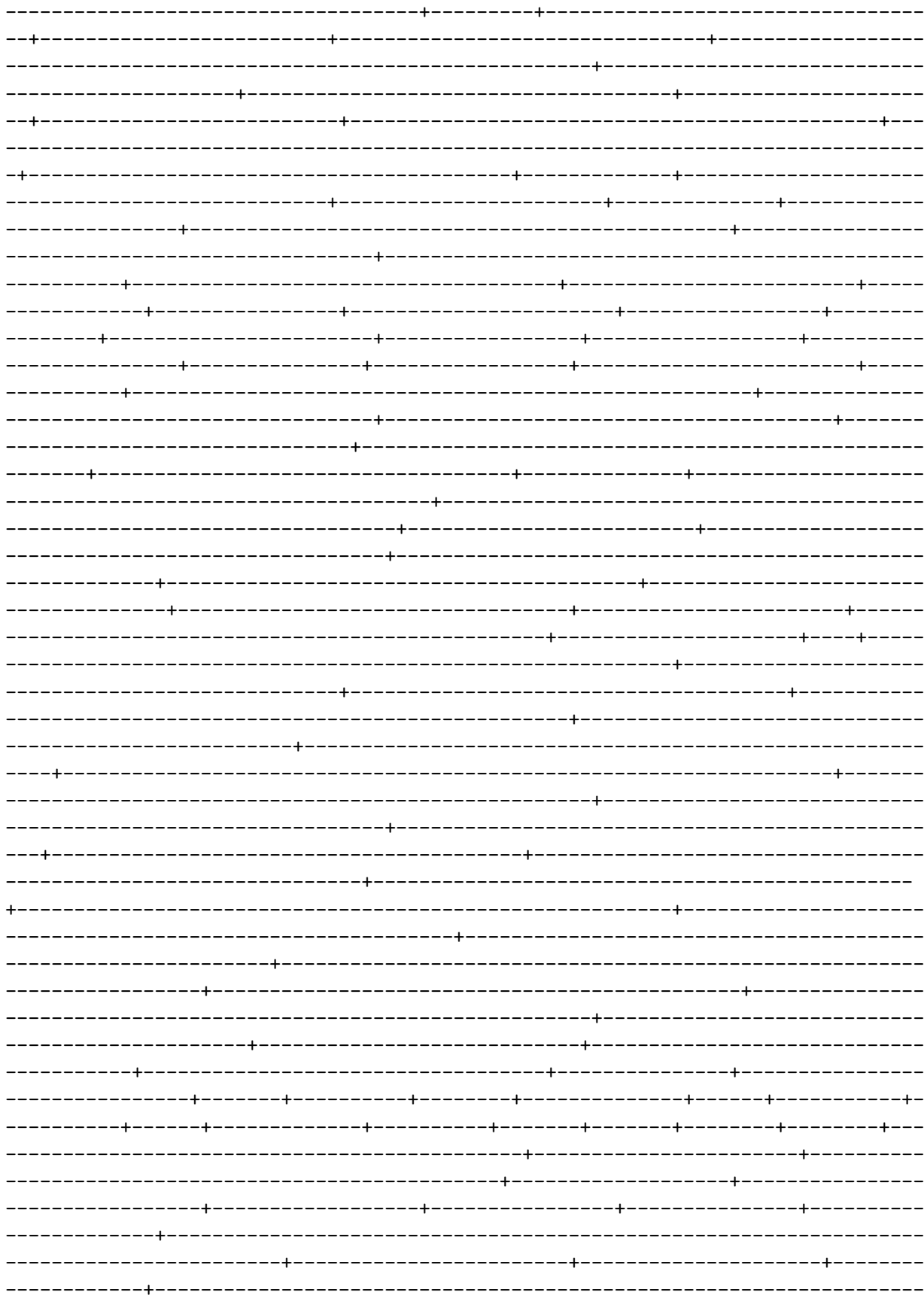
```
23/09/22 08:42:25 WARN DAGScheduler: Broadcasting large task binary with size
2.1 MiB
```

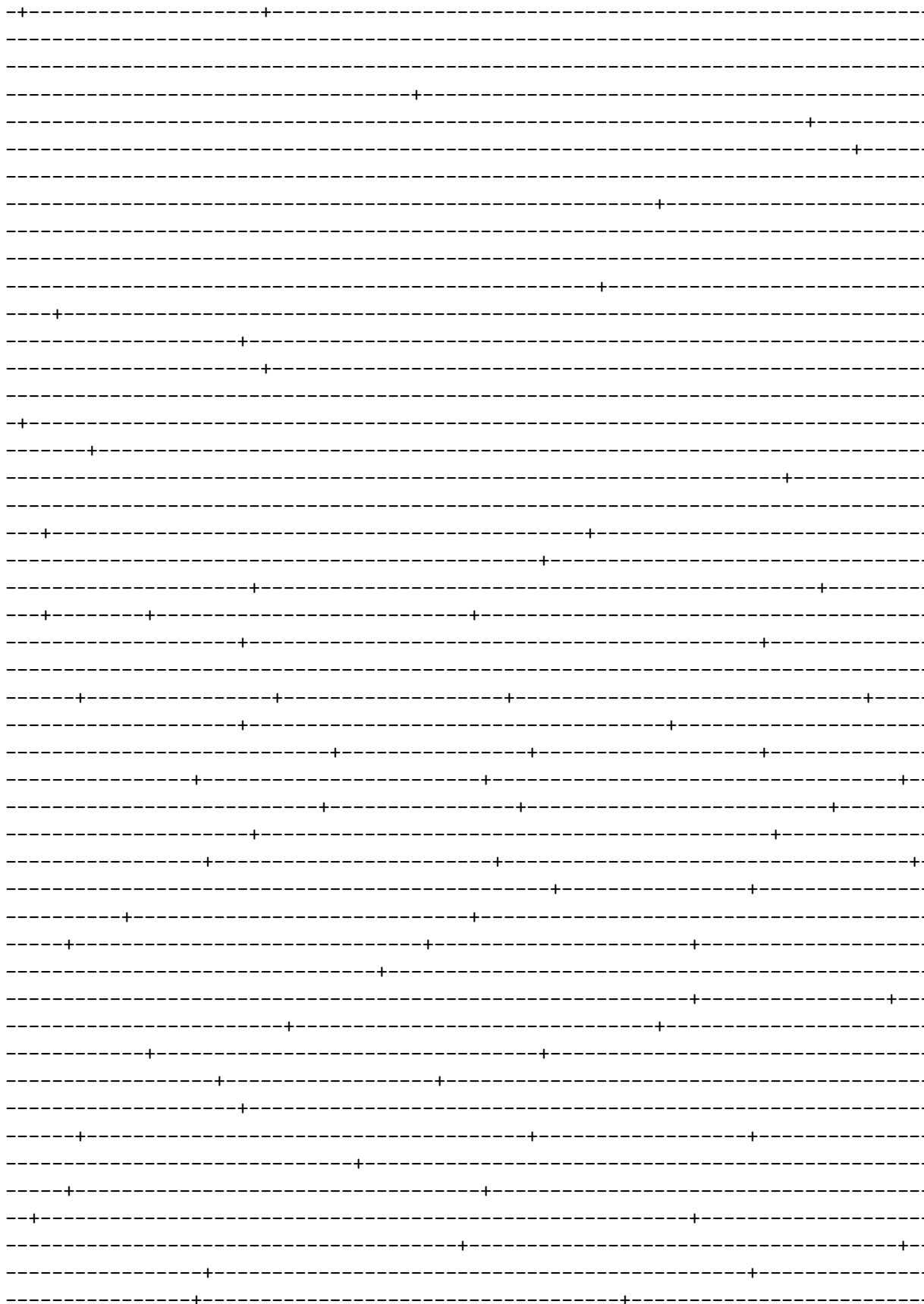
```
23/09/22 08:43:10 WARN DAGScheduler: Broadcasting large task binary with size
5.8 MiB
```

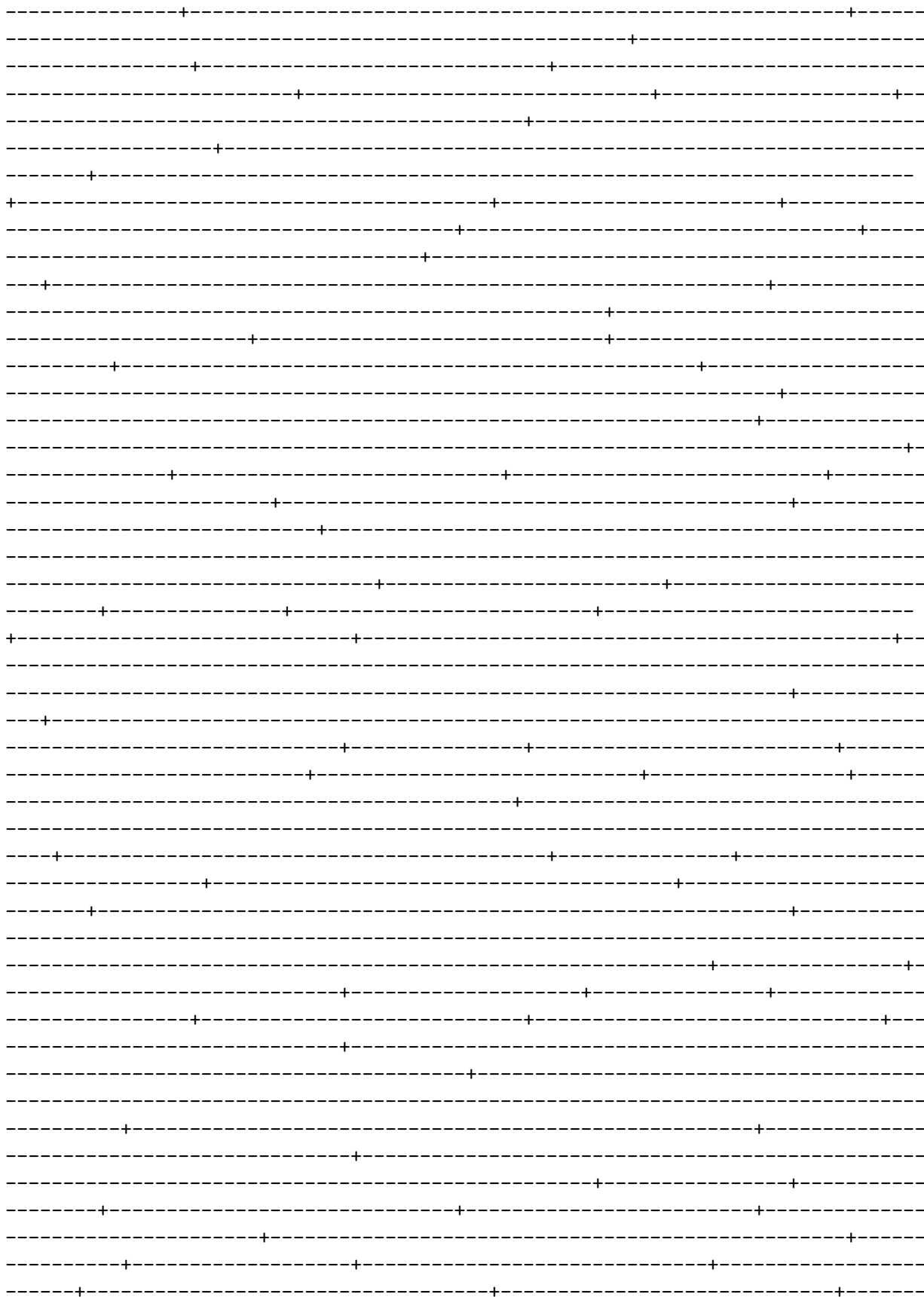


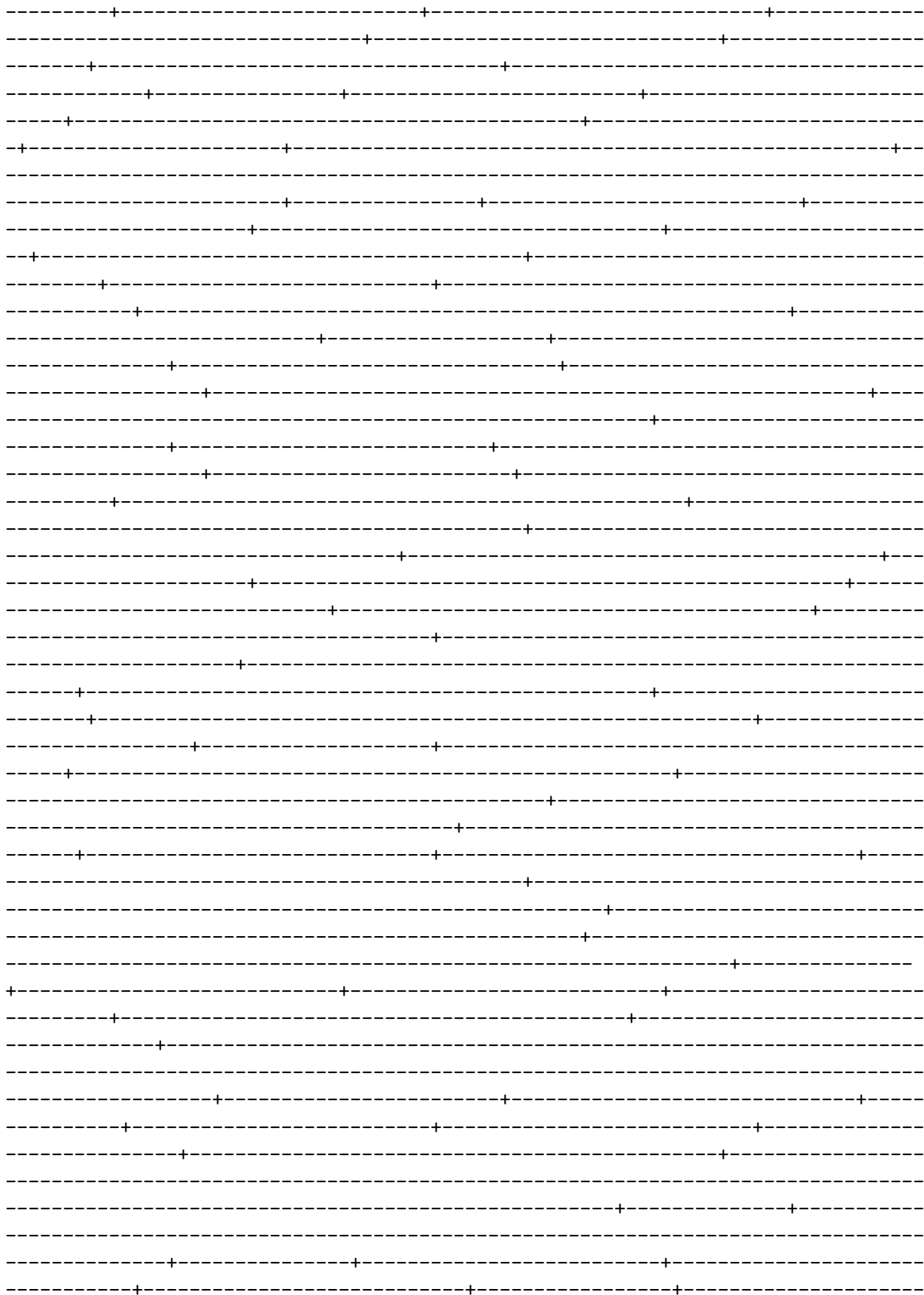


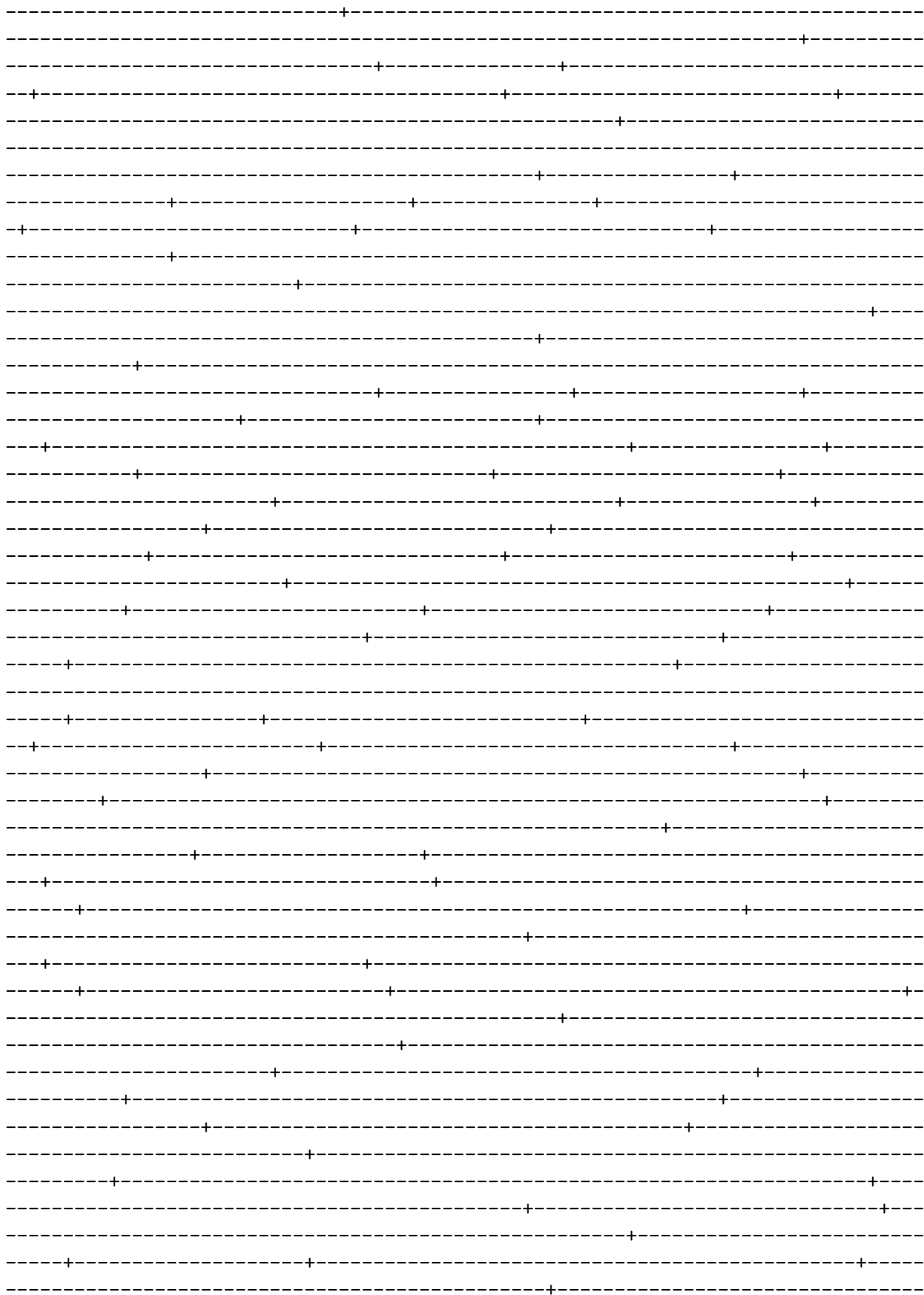


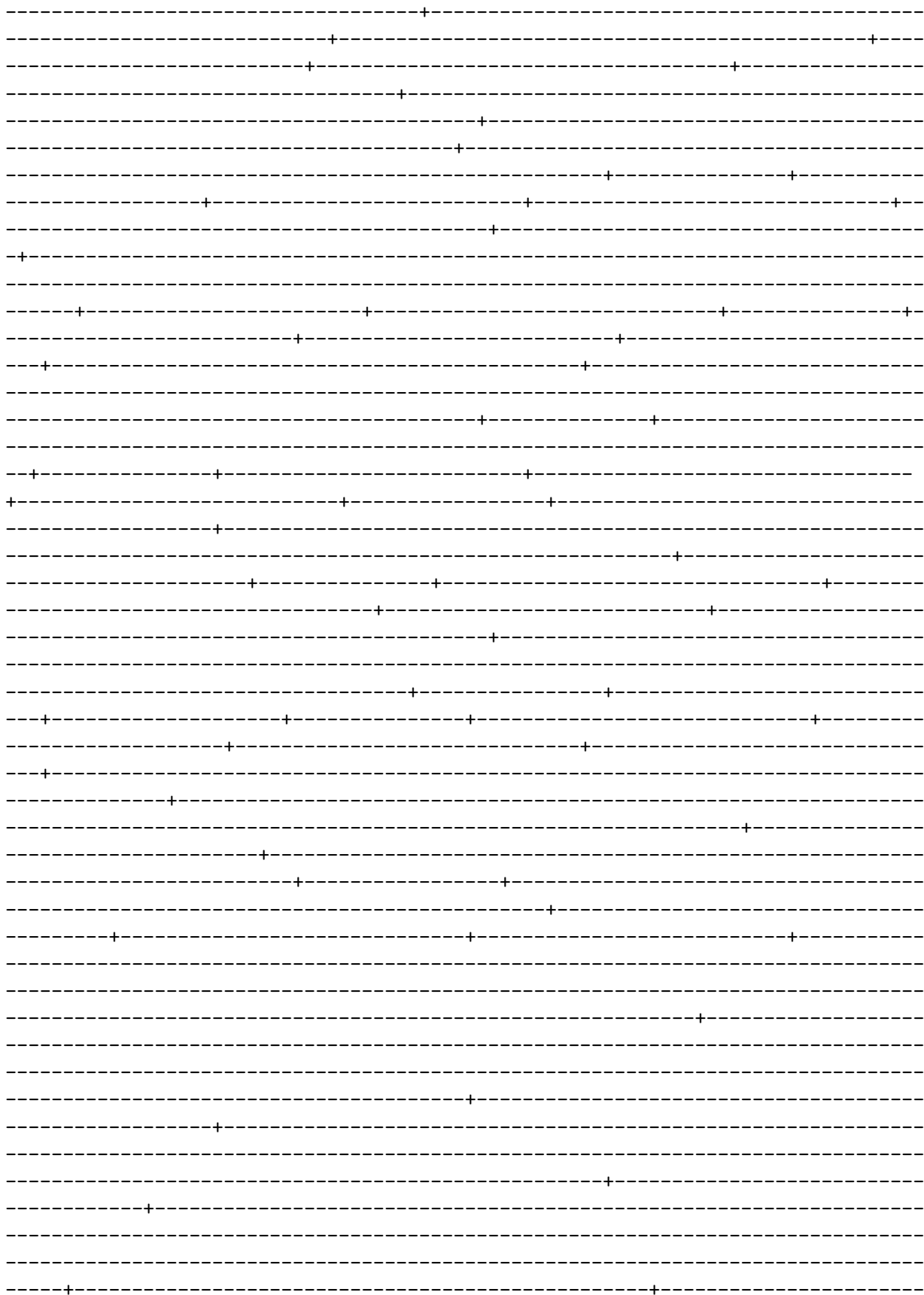












```

-----+-----
-----
-----+-----+-----+-----+-----
-----+-----+-----+-----+-----
-----+-----+-----+-----+-----
-----+-----+-----+-----+-----
-----+-----+-----+-----+-----
-----+-----+-----+-----+-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----
-----+
|summary|<!doctype html><html lang="en"><head> <title data-react-
helmet="true">data.world</title>    <meta charset="utf-8"><meta name="viewport"
content="width=device-width|initial-scale=1"><meta name="referrer"
content="origin-when-cross-origin"><meta name="og:image"
content="https://d2ogkqlrg66kh0.cloudfront.net/site-resources/images/fb-image-
share.3a87a6eb.png"><meta name="og:site_name" content="data.world"><meta
name="twitter:card" content="summary"><meta name="twitter:site"
content="@datadotworld"><meta property="fb:app_id"
content="776416075825233"><link rel="stylesheet" href="https://fonts.googleapis.
com/css2?family=Inconsolata:wght@400;700&family=Lato:ital|wght@0|300;0|400;0|700
;1|300;1|400;1|700&display=swap"><link rel="search" href="/opensearch.xml"
type="application/opensearchdescription+xml" title="data.world">  <link
rel="apple-touch-icon" sizes="57x57"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-57x57.png"><link rel="apple-touch-icon" sizes="60x60"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-60x60.png"><link rel="apple-touch-icon" sizes="72x72"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-72x72.png"><link rel="apple-touch-icon" sizes="76x76"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-76x76.png"><link rel="apple-touch-icon" sizes="114x114"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-114x114.png"><link rel="apple-touch-icon" sizes="120x120"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-120x120.png"><link rel="apple-touch-icon" sizes="144x144"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-144x144.png"><link rel="apple-touch-icon" sizes="152x152"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-152x152.png"><link rel="apple-touch-icon" sizes="180x180"

```

```

href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-180x180.png"><meta name="apple-mobile-web-app-capable"
content="yes"><meta name="apple-mobile-web-app-status-bar-style" content="black-
translucent"><meta name="apple-mobile-web-app-title" content="data.world"><meta
name="mobile-web-app-capable" content="yes"><meta name="theme-color"
content="#fff"><meta name="application-name" content="data.world"><link
rel="icon" type="image/png" sizes="32x32" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-32x32.png"><link rel="icon"
type="image/png" sizes="16x16" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-16x16.png"><link rel="shortcut icon" hr
ef="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon.ico"><link
href="https://assets.data.world/css/vendors~app.bf2ab82ca4e8e237d22a.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors~vendors.e418760f21d557e2e0e2.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors.5b2975e4af1441b468c5.css"
rel="stylesheet"><link
href="https://assets.data.world/css/app.1915b3c34db820707a63.css"
rel="stylesheet"></head><body><div id="root"><div class="page-container
CoreLayout__embed__1fxWq"><div class="view-container"><div
class="LoadingAnimation__loaderOverlay__2zSul
LoadingAnimation__default__KhFF6" data-dw="LoadingAnimation"><div
class="LoadingAnimation__loaderContainer__2arLY"><div class="LoadingAnimation__
spinner__2qQFp">Loading...</div></div></div></div></div></div></div><script
type="application/json" id="init-data">{"catalog":{"loadedResources":{}|"hoots":
{}|"clients":{}|"currentUser":{"profile":{"abTests":{"buckets":{"coreLayoutRefa
ctor":{"name":"inactive|"value":-1|"loadInstanceConfig":{"name":"load-
instance-config|"value":1|"sqlAlgebra":{"name":"preferLegacy|"value":2|"sqlQ
uery":{"name":"presto|"value":4}}|"accountStanding":"good|"activeSubscription
id":"","agentid":"","authorizationLevels":{"org":"READ|"allCatalogs":"NONE|"da
taLibrary":"NONE"}|"avatarUrl":"","bio":"","catalogStatistics":{"analysesCount":
0|"datasetsCount":0|"dataTypesCount":0|"glossaryCount":0|"projectsCount":0|"tabl
esCount":0|"displayName":"","email":"","created":"","company":"","emailVerified
":true|"location":"","numFollowers":0|"numFollowing":0|"numOrganizations":0|"org
Mode":false|"orgDetails":null|"orgPreferences":null|"relationship":{"member":fal
se|"followed":false|"staticGrants":{"analysisHubEnabled":false|"businessGlossar
yEnabled":false|"hootsEnabled":false|"lineageEnabled":false|"metadataCatalogEnab
led":false|"nativeLineageEnabled":false|"personalDatasetsDisabled":false|"preven
tGuestAuthorizations":false|"preventGuestUserInvite":false|"preventOpenDatasets"
:false|"preventSignedShareUrls":false|"samlAllowed":false|"thirdPartyAiEnabled":
false|"tags":[]|"updated":"","userCapabilities":{"canBulkEditAllMetadataResourc
es":false|"canCreateCollections":false|"canCreateDataResources":false|"canCreate
MetadataResources":false|"canDebugMetadataProfile":false|"canEditAllDatasetMetad
ata":false|"canManageAllCollectionAccess":false|"canManageAllCollectionHierarchy
":false|"canManageAllDataResources":false|"canManageAllHoots":false|"canManageOr
gAccess":false|"canManageOrgAutomations":false|"canManageOrgConnections":false|"
canManageOrgConnectionTasks":false|"canManageOrgSettings":false|"canManageOrgSub

```

```

scription":false|"canShowOrgAdminLabel":false|"hasOrgMembership":false|"visibil
ity":"","| "visitorid":"109156be-c4fb-41ea-b1b4-
efe1671c5808"| "website":"","| "authorizedAccounts":{"state":"idle"| "data":null}| "a
uthorizedClients":[]| "customSearch":{"default":{"customFacets":[]}| "byOrg":{"}}| "
dismissed":[]| "isLoggedIn":false|"needsAnalyticsAlias":false|"notifications":{"r
esults":[]| "unread":0| "unseen":0}| "onboardThisSession":false|"requests":[]| "requ
estsFetching":false|"tasks":null| "token":"eyJhbGciOiJIUzUxMiJ9.eyJzdWIiOiJwcm9kL
XVzZXItY2xpZW50IiwiaXNzIjoieYXV0aG9yaXR5OmRhdGFkb3R3b3JsZDo6MEJGMEVCRDMtMkRFNy00O
UZCLUi4N0ItQUQwMzA1N0JCODlFIiwiaWF0IjoxNjk1MzMxMjM1LCJyb2x1IjpbXSwiZXhwIjoxNjk1N
DE3NjM1LCJ2aXNpdG9yaWQiOiIxMDkxNTZiZS1jNGZiLTQxZWVtYjFiNC1lZmUxNjcxYzU4MDgiLCJnZ
W5lcmFsLXB1cnBvc2UiOnRydWUsImF1dGhvcml0eWlkcyI6WyJkYXRhZG90d29ybGQiXX0.nFFhdKNC
oIGQuVLekewk_RfKKtP4WTwb1raXlseSfGgyGhI6y5x1f5ZF180mOLpreLK7ySuqAHA88MFwOYP8AA"| "
unDismissedTourCount":2}| "dataset":{"hasError":false|"loadedDatasets":{"}| "usersD
atasets":{"}| "loadedCurrentUsersDatasets":false|"usersDatasetsTotalCount":0| "cust
omDataTypes":{"}| "workspacePreferences":{"}}| "discussion":{"}| "events":{"status":"c
losed"}| "integration":{"activeIntegrations":[]| "userJdbcSources":[]| "supportedJd
bcSources":[]}| "invite":{"}| "linkedOnboardModal":{"show":false}| "loginModal":{"sh
ow":false}| "queries":{"}| "relationships":{"}| "requestInfo":{"pathsMarkedAs404":[]|
"metadata":{"}| "requests":{"hasError":false|"loadedResources":{"}| "allPendingRequ
ests":{"count":0| "records":null| "hydrations":{"}}}| "routing":{"locationBeforeTran
sitions":null}| "signupModal":{"show":false|"treatment":null|"treatmentProps":nul
l}| "siteStatus":{"uploadsOperational":true|"queryOperational":true|"websiteOpera
tional":true|"downloadsOperational":true|"shouldAskReload":false|"shouldForceRel
oad":false|"connectionError":false|"incidentName":"","| "incidentMessage":"","| "pageS
tatus":"none"}| "treatments":{"isSite":false|"activeTreatments":["showVersions"|s
howWorkspaceTabs|showEntityLinkOptions|canCreateOpenDatasets|canCreateOrgs|showC
onversionPages|showSignup|showSocialLogin]}}| "user":{"hasError":false|"loadedUs
ers":{"}| "userFollows":{"}| "userLikes":{"}| "userRecentComments":{"}| "groupAuthorizat
ions":{"}| "featuredDatasets":{"loaded":false|"resources":[]}}}</script><script
type="application/json" id="init-config">{"adwords_pixel_id":"869213640"| "analyt
ics":{"segment_id":"IXhtZmPohuGM60VHk59cdSbWqWQBB7aR"| "segment_endpoint":"https:
\u002F\u002Fapi-el.data.world"}| "analytics_log_to_console":false|"base_cookie_do
main":null|"contact_admin_form_config":[{"type":"input"| "label":"Email"| "lookupP
roperty":"email"| "readOnly":true|162|"required":true}| {"type":"textarea"|164|"labe
l":"Message"|165|"placeholder":"Add a message (i.e details that need to be review
ed)"| "required":true}|167|"csrf_cookie_name":"_csrf"| "csrf_header_name":"x-csrf-
token"| "data_provider_agents":["johnsnowlabs"| "associatedpress"]| "download_conte
nt_url":"https:\u002F\u002Fdownload.dwcontent.com"| "download_url":"https:\u002F\u
002Fdownload.data.world"| "dw_tableau_wdc_host_domain":"https:\u002F\u002Ftablea
u.data.world"| "deployed_environment_full":"prod"| "facebook_pixel_id":"2619849941
77057"| "fb_app_id":"776416075825233"| "google_chooser_app_id":"956984791354"| "hel
p_cards":[{"id":"getting_started"| "header":"Get Started"| "body":"New to
data.world? Start here to find introductory
articles."| "link":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0
--Overview-of-data-world.html"| "imageSrc":"getting-started-
lightglow"}| {"id":"join_discussion"| "header":"Join the discussion"| "body":"Join
the general discussion or start your own
topic"| "link":"","| "imageSrc":"discussions-

```

```

lightglow"}|{"id":"see_documentation"| "header":"Visit the Help
Center"| "body":"Find general platform documentation and technical tutorials."|"l
ink":"https:\u002F\u002Fdocs.data.world\u002F"| "imageSrc":"documentation-
lightglow"}|{"id":"contact_support"| "header":"Contact support"| "body":"Report an
issue or make a
suggestion."|"link":"https:\u002F\u002Fhelp.data.world\u002Fhc\u002Fen-
us\u002Frequests\u002Fnew"| "imageSrc":"support-lightglow_sales-guy"}}|"hoot_url"
:""| "iframely_key":"e4239223dd4ad21531b3a044840d5898"| "is_monitored_env":true|"i
s_vpc":false|"max_upload_size":524288000|"media_upload_domain":"mediauploads.dat
a.world"| "media_url":"https:\u002F\u002Fmedia.data.world"| "oauth_authorize_url":
"https:\u002F\u002Fauth.data.world"| "oauth_providers":[{"name":"facebook"| "displ
ay_name":"Facebook"| "active":true}209|{"name":"github"| "display_name":"GitHub"|
"active":true}212|{"name":"google"| "display_name":"Google"| "active":true}215|{"na
me":"linkedin"| "display_name":"LinkedIn"| "active":false}|{"name":"twitter"| "disp
lay_name":"Twitter"| "active":true}]|"openAccessRoutePrefixes":["\u002Fplaineale
rprojects"| "\u002Fd1gi\u002Fmissing-fb-posts-w-share-
stats"| "\u002Ftravelwarnings\u002Ftravel-
danger"| "\u002Ffootprint\u002Fnfa-2018-edition"| "\u002Fpolymathic\u002Fcasualties-
of-the-syrian-civil-
war"| "\u002Faurielle\u002Ffortune-500-2017"| "\u002Fnjdotcom"| "\u002Fian\u002Fusa-baby-
names-national-and-state-by-state"| "\u002Fowentemple\u002Ftext-and-content-
features-of-most-persuasive-ted-talks"| "\u002Fschnee\u002Ftangled"| "\u002Fchhs\u002F
fc544658-35c5-4be0-af20-fc703bc57c13"| "\u002Fnz-stats-
nz\u002F4e5ebd0-4279-4292-a85f-33c9f2a93f8d"| "\u002Fshad\u002Fcovid-19-time-
series-data"| "\u002Fmarkmarkoh\u002Fcoronavirus-data"| "\u002Fdallemang\u002Fcovid-
19-data"| "\u002Fpolicinginamerica"| "\u002Findustryresearch\u002Fdata-engineer-
survey-2021"| "\u002Fembed\u002Fswwo|swwo"| "\u002Fassociatedpress\u002F2018-election-
results-precinct-level"| "\u002Faaron-kessler-ap\u002Fcongressional-
earmarks"| "\u002Fassociatedpress\u002Fap-
votecast"| "\u002Fassociatedpress\u002Fbolsa-familia-aid-to-brazils-
poor"| "\u002Fassociatedpress\u002Fcdc-covid-19-vaccine-
tracker"| "\u002Fassociatedpress\u002Fcovid-impact-survey-public-
data"| "\u002Fassociatedpress\u002Fcovid-tracking-project-testing-in-
states"| "\u002Fassociatedpress\u002Fap-planner-covid-reopening-
data"| "\u002Fassociatedpress\u002Ffema-distribution-of-ppe-to-
states"| "\u002Fassociatedpress\u002Fforeign-exchange-
rates"| "\u002Fassociatedpress\u002Fus-dams-
public"| "\u002Fassociatedpress\u002Fjohns-hopkins-coronavirus-case-
tracker"| "\u002Fassociatedpress\u002Fmass-killings-
public"| "\u002Fassociatedpress\u002Fopioid-
prescriptions-2010-2015"| "\u002Fassociatedpress\u002Fpublic-health-official-
departures"| "\u002Fassociatedpress\u002Frefugee-admissions-to-us-end-
fy-2018"| "\u002Fassociatedpress\u002Frental-assistance-from-coronavirus-relief-
fund"| "\u002Fassociatedpress\u002Fstate-ppe-
purchases"| "\u002Fassociatedpress\u002Fmarshall-project-covid-cases-in-
prisons"| "\u002Fassociatedpress\u002Fvehicle-miles-traveled"| "\u002Ffrare\u002Ffish-
surveys"| "\u002Ffrare\u002Flarval-dispersal-data"]|"preferred_oauth_provider":"goo
gle"| "prod_env":true|"profile_interests":["education"| finance| government| busines

```


s|transportation|health|agriculture|environment|sports|entertainment|technology|science|culture|politics|crime"]|"query_url":"https:\u002F\u002Fquery.data.world|"regional_cluster":"US"|"request_dataset_access_form_config":[{"type":"input"|"label":"Full name"|"lookupProperty":"displayName"284|"required":true285|"readOnly":true}]{ "type":"email"|"label":"Contact email"|"placeholder":"Provide an email in the case a dataset administrator needs to reach out"289|"showOptionalFlag":true}]{ "type":"textarea"291|"label":"Message"292|"placeholder":"Share what you would like to do with this dataset"293|"required":true}]294|"requests":{"contributionSuggestionInfoText":"You are suggesting changes to this \u003C%= resourceType %\u003E that will need to be approved by an admin of this organization. Only you and admins will be able to see your changes until they're approved."|"pendingApprovalsText":"There \u003C%= pendingApprovalsNumber %\u003E from the community pending admin approval."|"pendingSuggestionsText":"Your suggested changes are visible on the Suggestions tab| but only you and this organization's admins can see them until they have been approved. The Overview tab displays only approved information."|"personalResourceSharedWithText":"This draft \u003C%= resourceType %\u003E is within your private collection and only visible to you. It is neither discoverable nor viewable by the data.world community. To open it up to others to view| publish it to an organization."|"transferOwnershipMessage":"By publishing this \u003C%= resourceType %\u003E to an organization| you are transferring ownership to the organization. Once your draft is published| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."|"transferOwnershipModalWarningBanner":"Once the organization has published your draft| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."}|"sentry_dsn_client":"https:\u002F\u002Ff763f0cf14f04c75b4789a118ab1beb3@sentry.io\u002F1486857"|"server_origin":"https:\u002F\u002Fdata.world"|"sparql_host":"https:\u002F\u002Fpdw-sparqlproxy-us-east-1.prod.data.world"|"stripe_public_key":"pk_live_CPGU5gVFxlkY6pabDXladsMR"|"swagger_use_case_label_override_org":["oa-test"|engie-impact|"e-ddw"]|"tableau_api_version":"2.8"|"ui":{"orgAdmin":{"resourceLinks":[{"key":"getting-started"|"linkDisplayName":"Getting started guide"315|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"316|"newTab":true317|"svgIcon":"meta"318|"verticallyAlignIcon":true}319|{"key":"documentation"320|"linkDisplayName":"Documentation"321|"url":"https:\u002F\u002Fdocs.data.world\u002F"322|"newTab":true323|"svgIcon":"meta"324|"verticallyAlignIcon":true}325|{"key":"integrations"326|"linkDisplayName":"Integrations"327|"url":"\u002Fintegrations"328|"newTab":true329|"svgIcon":"integration"330|"verticallyAlignIcon":true331|"showIfAnyDismissed":["onboard-checklist"332|"onboard-checklist-admin"]}333|{"key":"new-features"334|"linkDisplayName":"New features"335|"url":"https:\u002F\u002Fwhatsnew.data.world"336|"newTab":true337|"svgIcon":"magic-wand"338|"verticallyAlignIcon":true339|"showIfAnyDismissed":["onboard-checklist"340|"onboard-checklist-admin"]}341|{"key":"enterprise"342|"linkDisplayName":"data.world for the enterprise"343|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-demo\u002F?src=inapp_data.world-for-the-enterprise"344|"newTab":true345|"svgIcon

```

": "organization"346| "verticallyAlignIcon": true}]347| "onboardChecklist": {"key": "site"348| "items": [{"id": "createAccount"349| "analyticsId": "onboard_checklist.create_account"350| "complete": true351| "formattedMessage": {"id": "onboard_checklist.create_account"352| "description": "Text for onboard checklist create account link"353| "defaultMessage": "Create account"}}354| {"id": "inviteTeam"|"analyticsId": "onboard_checklist.invite-team"|"href": "\u002F\u003C%- orgAgentid %\u003E?tab=members"|"dismissOnClick": "invite-team-link"|"completeIfDismissed": "invite-team-link"|"formattedMessage": {"id": "onboard_checklist.invite-team"|"description": "Text for onboard checklist invite team link"|"defaultMessage": "Invite your team"}}| {"id": "connectData"|"analyticsId": "onboard_checklist.connect-data"|"href": "\u002Fcreate-a-dataset"|"dismissOnClick": "connect-data-link"|"completeIfDismissed": "connect-data-link"|"formattedMessage": {"id": "onboard_checklist.connect-data"|"description": "Text for onboard checklist connect data link"|"defaultMessage": "Connect data sources"}}| {"id": "exploreIntegrations"371| "analyticsId": "onboard_checklist.explore-integrations"372| "href": "\u002Fintegrations"373| "target": "_blank"374| "dismissOnClick": "explore-integrations-link"375| "completeIfDismissed": "explore-integrations-link"376| "formattedMessage": {"id": "onboard_checklist.explore-integrations"377| "description": "Text for onboard checklist explore integrations link"378| "defaultMessage": "Explore integrations"}}| {"id": "createProject"380| "analyticsId": "onboard_checklist.create-project"381| "href": "\u002Fcreate-a-project"382| "dismissOnClick": "create-project-link"383| "completeIfDismissed": "create-project-link"384| "formattedMessage": {"id": "onboard_checklist.create-project"385| "description": "Text for onboard checklist create project link"386| "defaultMessage": "Create your first project"}}}]| "onboardTour": [{"show": true388| "header": "Welcome to data.world!"389| "subhead": "The place for modern data teamwork."390| "leftCol": "Organize data work so you and your team find what you need&nbsp;faster."391| "middleCol": "Explore and create a shared understanding of your&nbsp;data."392| "rightCol": "Share insights and invite more people to participate in data&nbsp;teamwork."}393| {"show": true394| "imageHeaderKey": "search"395| "imageKey": "collaborate"396| "imageAlt": "Start with Data"397| "header": "Start with fantastic&nbsp;data"398| "subhead": "Browse thousands of free399| open data resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from almost anywhere into a single searchable&nbsp;library.\n\nMake it open400| share with your team401| or keep it&nbsp;private."}402| {"show": true403| "imageHeaderKey": "bulb"404| "imageKey": "dataProject"405| "imageAlt": "Data Projects"406| "header": "Organize your data&nbsp;projects"407| "subhead": "Create projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd context to make it easier for others to discover408| understand409| and contribute to your&nbsp;projects.\n\nImmediately query data without leaving your&nbsp;projects."}410| {"show": true411| "imageHeaderKey": "code"412| "imageKey": "slideIntegrations"413| "imageAlt": "Integrations"414| "reverse": true415| "header": "Conne

```

ct the data tools you love"416|"subhead":"Automatically keep data.world synced with your data sources.\n\nIntegrate your team's favorite tools417| from analysis to visualization."}418|{"show":true419|"imageHeaderKey":"discussions"420|"imageKey":"discussionsCollaboration"421|"imageAlt":"Contributors"422|"header":"Add more brainpower to your projects"423|"subhead":"Invite your team to securely and effectively work together.\n\nManage fine-grained permissions for each member.\n\nAsk questions424| add insights425| bookmark top contributions426| and more."}427|{"show":true428|"imageHeaderKey":"insight"429|"imageKey":"knowYourData"430|"imageAlt":"know your data"431|"header":"Share your results"432|"subhead":"Share meaningful insights and the data behind them.\n\nDistribute clean433| documented data for others to use in their own work.\n\nHelp the right people discover the resources they need."}}434|"orgMember":{"resourceLinks":[{"key":"getting-started|"linkDisplayName":"Getting started guide"436|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"437|"newTab":true438|"svgIcon":"meta"439|"verticallyAlignIcon":true}440|{"key":"documentation"441|"linkDisplayName":"Documentation"442|"url":"https:\u002F\u002Fdocs.data.world\u002F"443|"newTab":true444|"svgIcon":"meta"445|"verticallyAlignIcon":true}446|{"key":"integrations"447|"linkDisplayName":"Integrations"448|"url":"\u002Fintegrations"449|"newTab":true450|"svgIcon":"integration"451|"verticallyAlignIcon":true452|"showIfAnyDismissed":["onboard-checklist"453|"onboard-checklist-admin"]}454|{"key":"new-features"455|"linkDisplayName":"New features"456|"url":"https:\u002F\u002Fwhatsnew.data.world"457|"newTab":true458|"svgIcon":"magic-wand"459|"verticallyAlignIcon":true460|"showIfAnyDismissed":["onboard-checklist"461|"onboard-checklist-admin"]}462|{"key":"enterprise"463|"linkDisplayName":"data.world for the enterprise"464|"url":"https:\u002F\u002Fdata.world\u002Fschedula-demo\u002Fsrc=inapp_data.world-for-the-enterprise"465|"newTab":true466|"svgIcon":"organization"467|"verticallyAlignIcon":true}}468|"onboardChecklist":{"key":"site"469|"items":[{"id":"joinOrg|"analyticsId":"onboard_checklist.join-org|"href":"\u002Fsettings\u002Forganizations|"completeIfMemberOfOrg":true|"formattedMessage":{"id":"onboard_checklist.join-org|"description":"Text for onboard checklist join org link|"defaultMessage":"Join organization"}}|{"id":"takeTour"477|"analyticsId":"onboard_checklist.take_tour"478|"showOnboardModalOnClick":true479|"completeIfDismissed":"onboard-tour-complete"480|"formattedMessage":{"id":"onboard_checklist.take_tour"481|"description":"Text for onboard checklist take tour link"482|"defaultMessage":"Take the tour"}}483|{"id":"customizeProfile"484|"analyticsId":"onboard_checklist.customize_profile"485|"href":"\u002Fsettings"486|"dismissOnClick":"customize-profile-link"487|"completeIfDismissed":"customize-profile-link"488|"formattedMessage":{"id":"onboard_checklist.customize_profile"489|"description":"Text for onboard checklist customize profile link"490|"defaultMessage":"Upload your photo"}}491|{"id":"createProject"492|"analyticsId":"onboard_checklist.create-project"493|"href":"\u002Fcreate-a-project"494|"dismissOnClick":"create-project-link"495|"completeIfDismissed":"create-project-link"496|"formattedMessage":{"id":"onboard_checklist.create-project"497|"description":"Text for onboard checklist create project

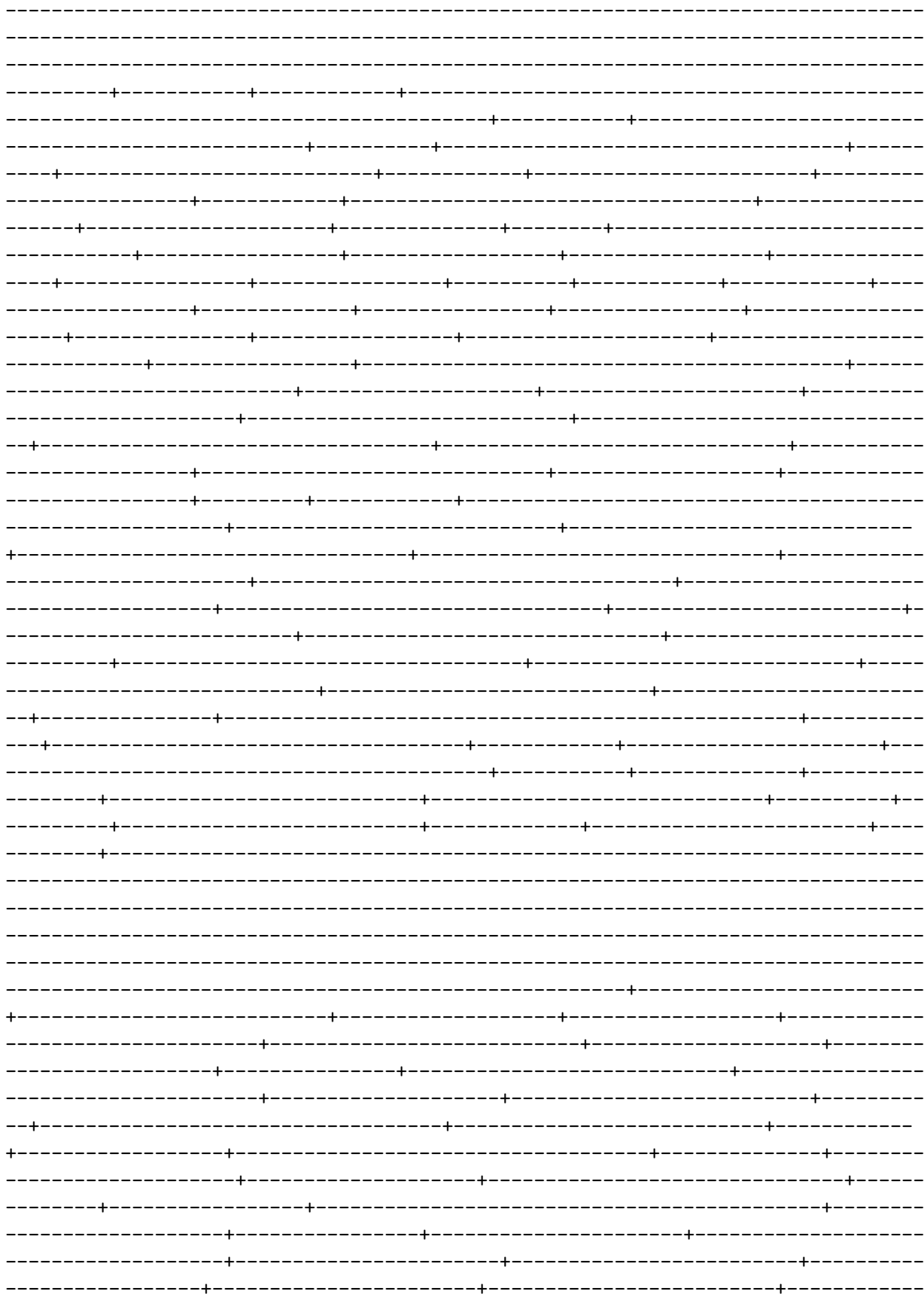
link"498|"defaultMessage":"Create your first project"}}499|{"id":"exploreIntegrations"500|"analyticsId":"onboard_checklist.explore-integrations"501|"href":"\u002Fintegrations"502|"target":"_blank"503|"dismissOnClick":"explore-integrations-link"504|"completeIfDismissed":"explore-integrations-link"505|"formattedMessage":{"id":"onboard_checklist.explore-integrations"506|"description":"Text for onboard checklist explore integrations link"507|"defaultMessage":"Explore integrations"}}}}508|"onboardTour":[{"show":true509|"header":"Welcome to data.world!"510|"subhead":"The place for modern data teamwork."511|"leftCol":"Organize data work so you and your team find what you need faster."512|"middleCol":"Explore and create a shared understanding of your data."513|"rightCol":"Share insights and invite more people to participate in data teamwork."}514|{"show":true515|"imageHeaderKey":"search"516|"imageKey":"collaborate"517|"imageAlt":"Start with Data"518|"header":"Start with fantastic data"519|"subhead":"Browse thousands of free520| open data resources; bookmark your favorites for later.\n\nGather your own data from almost anywhere into a single searchable library.\n\nMake it open521| share with your team522| or keep it private."}523|{"show":true524|"imageHeaderKey":"bulb"525|"imageKey":"dataProject"526|"imageAlt":"Data Projects"527|"header":"Organize your data projects"528|"subhead":"Create projects that hold everything you need to answer critical questions.\n\nAdd context to make it easier for others to discover529| understand530| and contribute to your projects.\n\nImmediately query data without leaving your projects."}531|{"show":true532|"imageHeaderKey":"code"533|"imageKey":"slideIntegrations"534|"imageAlt":"Integrations"535|"reverse":true536|"header":"Connect the data tools you love"537|"subhead":"Automatically keep data.world synced with your data sources.\n\nIntegrate your team's favorite tools538| from analysis to visualization."}539|{"show":true540|"imageHeaderKey":"discussions"541|"imageKey":"discussionsCollaboration"542|"imageAlt":"Contributors"543|"header":"Add more brainpower to your projects"544|"subhead":"Invite your team to securely and effectively work together.\n\nManage fine-grained permissions for each member.\n\nAsk questions545| add insights546| bookmark top contributions547| and more."}548|{"show":true549|"imageHeaderKey":"insight"550|"imageKey":"knowYourData"551|"imageAlt":"know your data"552|"header":"Share your results"553|"subhead":"Share meaningful insights and the data behind them.\n\nDistribute clean554| documented data for others to use in their own work.\n\nHelp the right people discover the resources they need."}]]555|"other":{"resourceLinks":[{"key":"getting-started|"linkDisplayName":"Getting started guide"557|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"558|"newTab":true559|"svgIcon":"meta"560|"verticallyAlignIcon":true}561|{"key":"documentation"562|"linkDisplayName":"Documentation"563|"url":"https:\u002F\u002Fdocs.data.world\u002F"564|"newTab":true565|"svgIcon":"meta"566|"verticallyAlignIcon":true}567|{"key":"integrations"568|"linkDisplayName":"Integrations"569|"url":"\u002Fintegrations"570|"newTab":true571|"svgIcon":"integration"572|"verticallyAlignIcon":true573|"showIfAnyDismissed":["onboard-checklist"574|"onboard-checklist-admin"]}575|{"key":"new-features"576|"linkDisplayName":"New features"577|"url":"https:\u002F\u002Fwhatsn

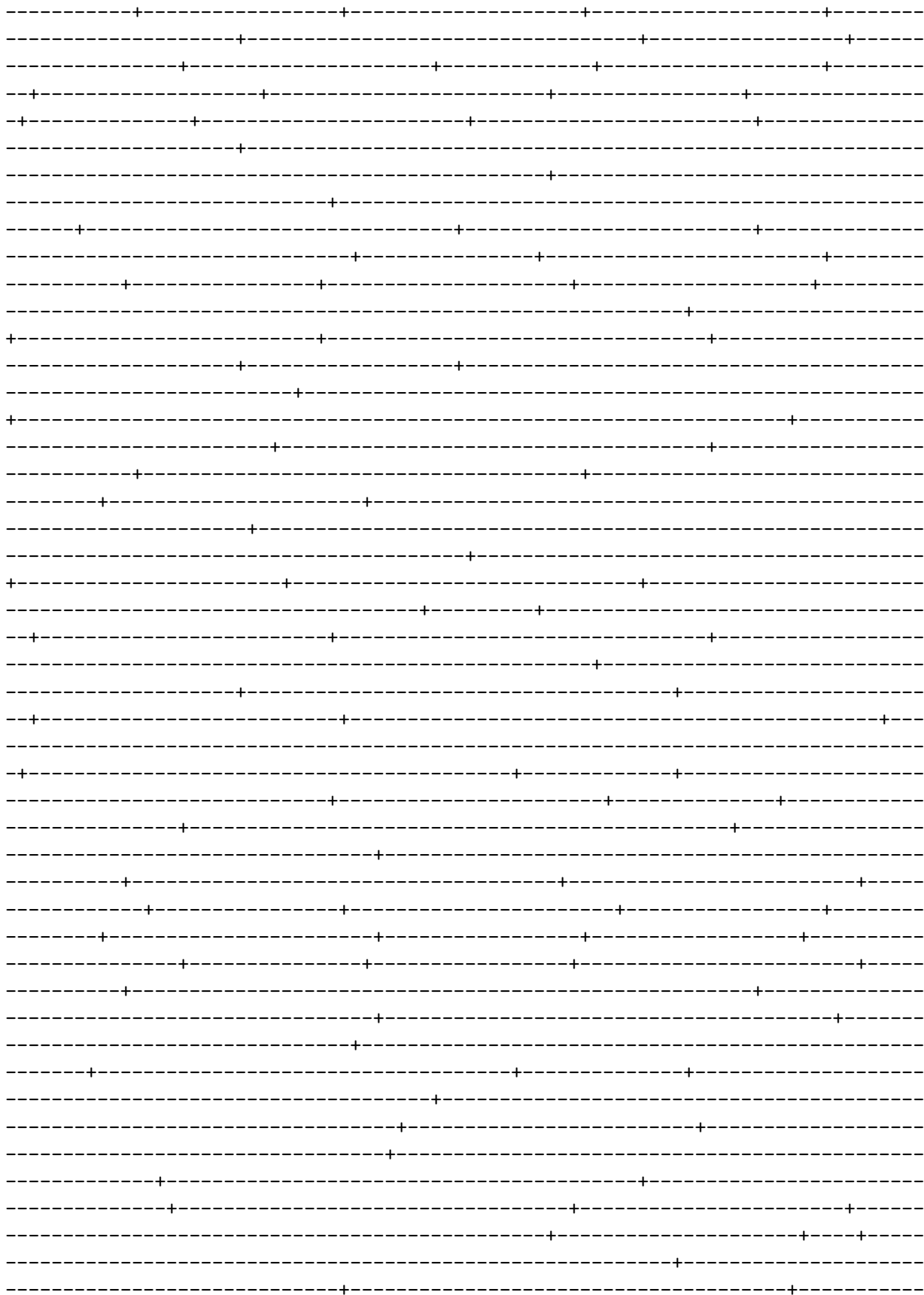
```

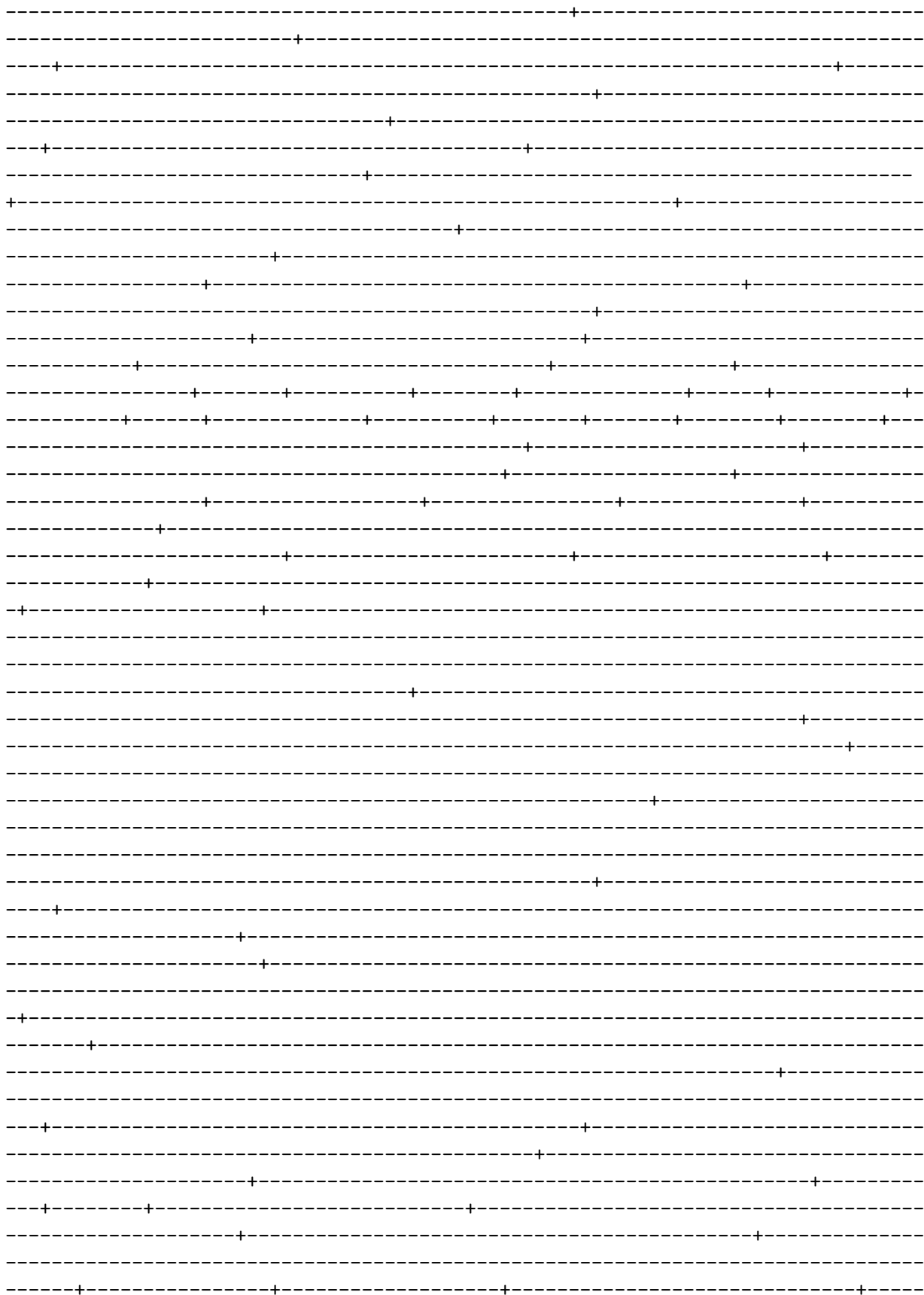
ew.data.world"578|"newTab":true579|"svgIcon":"magic-
wand"580|"verticallyAlignIcon":true581|"showIfAnyDismissed":["onboard-
checklist"582|"onboard-checklist-
admin"]}]583|{"key":"enterprise"584|"linkDisplayName":"data.world for the
enterprise"585|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-
demo\u002F?src=inapp_data.world-for-the-enterprise"586|"newTab":true587|"svgIcon
":"organization"588|"verticallyAlignIcon":true}]589|"enableImageUpload":true|"on
boardChecklist":{"key":"site"591|"items":[{"id":"createAccount"592|"analyticsId
":"onboard_checklist.create_account"593|"complete":true594|"formattedMessage":{"i
d":"onboard_checklist.create_account"595|"description":"Text for onboard
checklist create account link"596|"defaultMessage":"Create account"}}597|{"id":"
takeTour"598|"analyticsId":"onboard_checklist.take_tour"599|"showOnboardModalOnC
lick":true600|"completeIfDismissed":"onboard-tour-complete"601|"formattedMessage
":{"id":"onboard_checklist.take_tour"602|"description":"Text for onboard
checklist take tour link"603|"defaultMessage":"Take the tour"}}604|{"id":"custom
izeProfile"605|"analyticsId":"onboard_checklist.customize_profile"606|"href":"\u
002Fsettings"607|"dismissOnClick":"customize-profile-
link"608|"completeIfDismissed":"customize-profile-link"609|"formattedMessage":{"
id":"onboard_checklist.customize_profile"610|"description":"Text for onboard
checklist customize profile link"611|"defaultMessage":"Upload your
photo"}}612|{"id":"createProject"613|"analyticsId":"onboard_checklist.create-
project"614|"href":"\u002Fcreate-a-project"615|"dismissOnClick":"create-project-
link"616|"completeIfDismissed":"create-project-
link"617|"formattedMessage":{"id":"onboard_checklist.create-
project"618|"description":"Text for onboard checklist create project
link"619|"defaultMessage":"Create your first project"}}620|{"id":"exploreIntegra
tions"621|"analyticsId":"onboard_checklist.explore-integrations"622|"href":"\u00
2Fintegrations"623|"target":"_blank"624|"dismissOnClick":"explore-integrations-
link"625|"completeIfDismissed":"explore-integrations-
link"626|"formattedMessage":{"id":"onboard_checklist.explore-
integrations"627|"description":"Text for onboard checklist explore integrations
link"628|"defaultMessage":"Explore
integrations"}}}]629|"onboardTour":[{"show":true630|"header":"Welcome to
data.world!"631|"subhead":"The place for modern data
teamwork."632|"leftCol":"Organize data work so you and your team find what you
need&nbsp;faster."633|"middleCol":"Explore and create a shared understanding of
your&nbsp;data."634|"rightCol":"Share insights and invite more people to
participate in data&nbsp;teamwork."}635|{"show":true636|"imageHeaderKey":"search
"637|"imageKey":"collaborate"638|"imageAlt":"Start with Data"639|"header":"Start
with fantastic&nbsp;data"640|"subhead":"Browse thousands of free641| open data
resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from
almost anywhere into a single searchable&nbsp;library.\n\nMake it open642| share
with your team643| or keep it&nbsp;private."}644|{"show":true645|"imageHeaderKey
":"bulb"646|"imageKey":"dataProject"647|"imageAlt":"Data
Projects"648|"header":"Organize your data&nbsp;projects"649|"subhead":"Create
projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd
context to make it easier for others to discover650| understand651| and
contribute to your&nbsp;projects.\n\nImmediately query data without leaving your

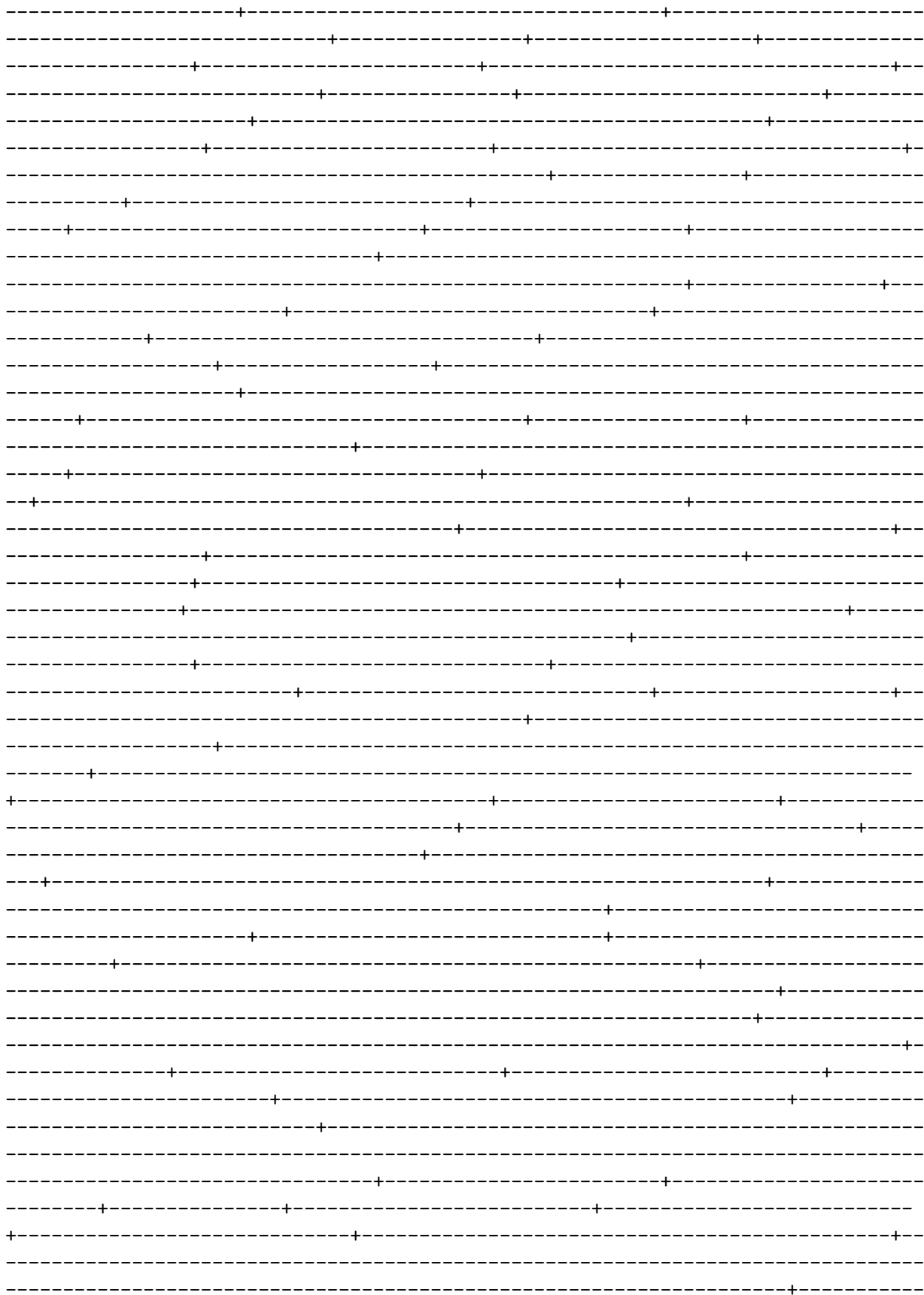
```

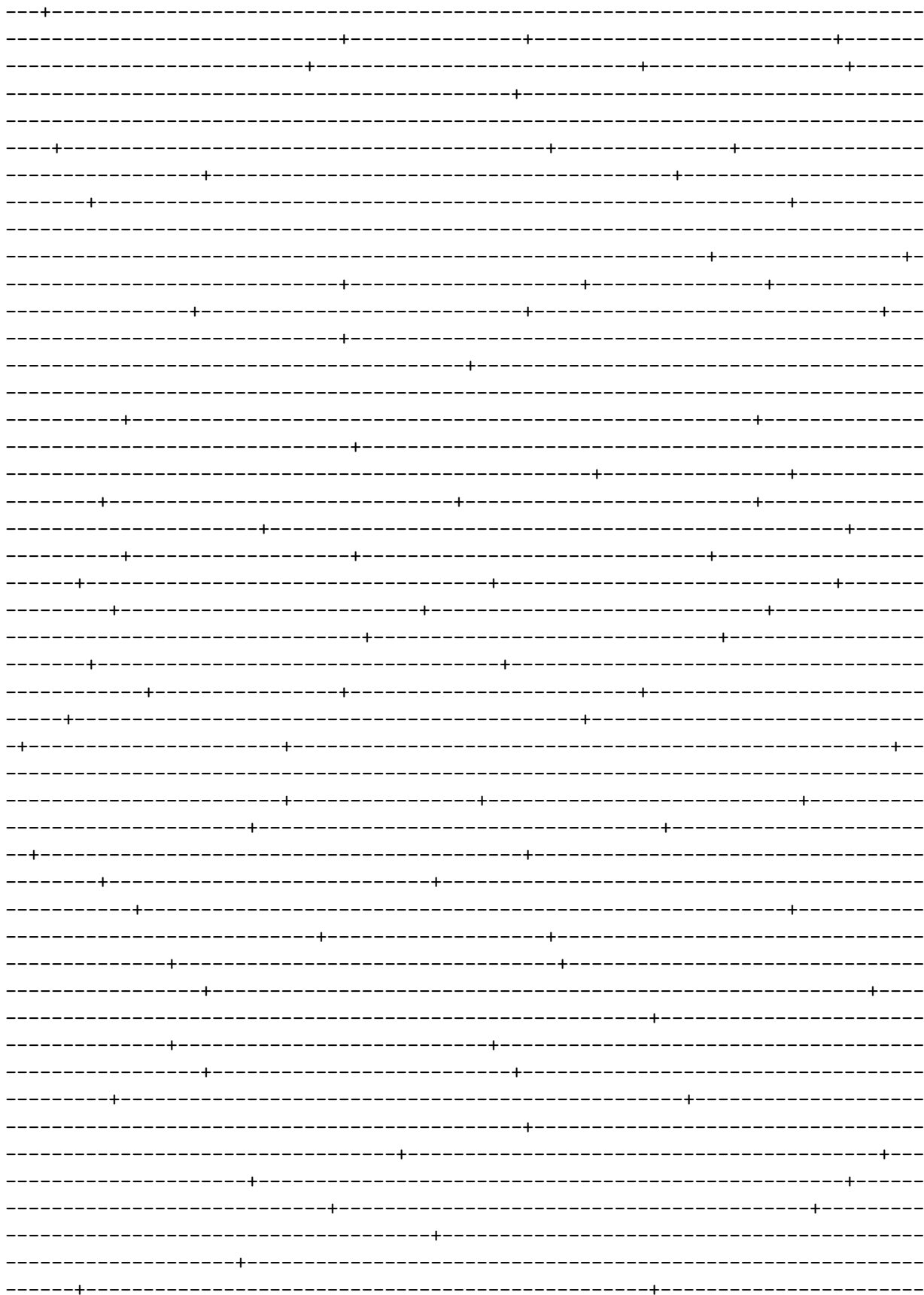
 projects."}652|{"show":true653|"imageHeaderKey":"code"654|"imageKey":"slid
 eIntegrations"655|"imageAlt":"Integrations"656|"reverse":true657|"header":"Conne
 ct the data tools you love"658|"subhead":"Automatically keep data.world
 synced with your data sources.\n\nIntegrate your team's favorite tools659|
 from analysis to visualization."}660|{"show":true661|"imageHeaderKey":"disc
 ussions"662|"imageKey":"discussionsCollaboration"663|"imageAlt":"Contributors"66
 4|"header":"Add more brainpower to your projects"665|"subhead":"Invite
 your team to securely and effectively work together.\n\nManage fine-grained
 permissions for each member.\n\nAsk questions666| add insights667| bookmark
 top contributions668| and more."}669|{"show":true670|"imageHeaderKey":"insi
 ght"671|"imageKey":"knowYourData"672|"imageAlt":"know your
 data"673|"header":"Share your results"674|"subhead":"Share meaningful insights
 and the data behind them.\n\nDistribute clean675| documented data for
 others to use in their own work.\n\nHelp the right people discover the
 resources they need."}]676|"newResourceItems":[{"type":"dataset|"exampleL
 ink":"https:\u002F\u002Fdata.world\u002F\ninja\u002Fanw-obstacle-history"}|{"type
 ":"project|"exampleLink":"https:\u002F\u002Fdata.world\u002Fmarkmarkoh\u002Fexp
 loring-thor"}]|"version_build":"20230921123426_7d6ba6c4|"version_check_interva
 l":60000|"version_commit":"7d6ba6c4|"visibility":{"discoverableVisibilitySubHea
 derText":"This \u003C%= resourceType %\u003E can be discovered by all data.world
 users. You can change the \u003C%= resourceType %\u003E visibility or make it
 private in the table below."|"openVisibilitySubHeaderText":"This \u003C%=
 resourceType %\u003E can be discovered and viewed by anyone on data.world. You
 can change the \u003C%= resourceType %\u003E visibility or make it private in
 the table below."|"privateVisibilityShareButtonText":"Share with
 community"|"privateVisibilitySubHeaderText":"This \u003C%= resourceType %\u003E
 is only visible to you. It is neither discoverable nor viewable by the
 data.world community. To open it up to others to view| change the visibility
 settings below."|"shareModalBodyText":"Set the visibility level for all of
 data.world below. The content of the \u003C%= resourceType %\u003E will only be
 visible to those who have been granted explicit access to this \u003C%=
 resourceType %\u003E."|"shareModalHeaderText":"Share with all
 data.world"|"visibilityTableTitleCellText":"Anyone on
 data.world"}|"context_id":"b04792da5b608cdd"}</script><script
 type="application/json" id="init-profile">{"requestAccessForm":[{"label":"Full n
 ame"|"required":true693|"readOnly":true694|"showOptionalFlag":false695|"lookupPr
 operty":"displayName"696|"type":"input"}|{"label":"Contact
 email"|"placeholder":"Provide an email in the case a dataset administrator needs
 to reach out"699|"required":false|"readOnly":false701|"showOptionalFlag":true|"t
 ype":"email"}|{"label":"Message"|"placeholder":"Share what you would like to do
 with this dataset"705|"required":true706|"readOnly":false707|"showOptionalFlag":
 false708|"type":"textarea"}]}</script><script type="application/json" id="init-
 analytics">{"siteId":"00000"}</script><script src="https://js.stripe.com/v3/"
 async id="stripe-js"></script> <script type="text/javascript" src="https://asse
 ts.data.world/js/runtime.4c0777f218351a402063.js"></script><script
 type="text/javascript" src="https://assets.data.world/js/vendors~app.73c77cf6388
 8ff69bcb2.js"></script><script type="text/javascript" src="https://assets.data.w
 orld/js/vendors~app~vendors.66b69a1400c3cdba7c1a.js"></script><script

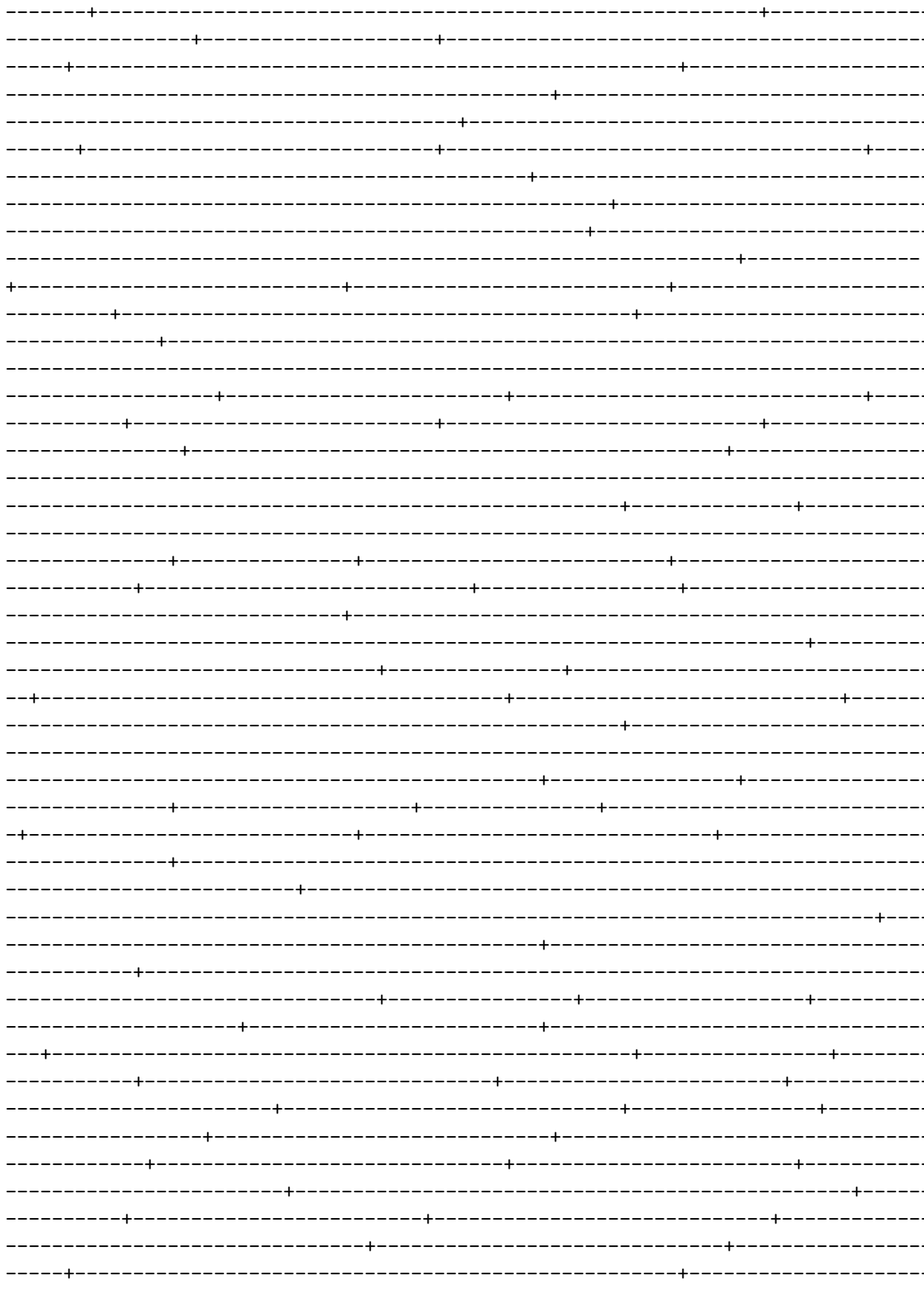


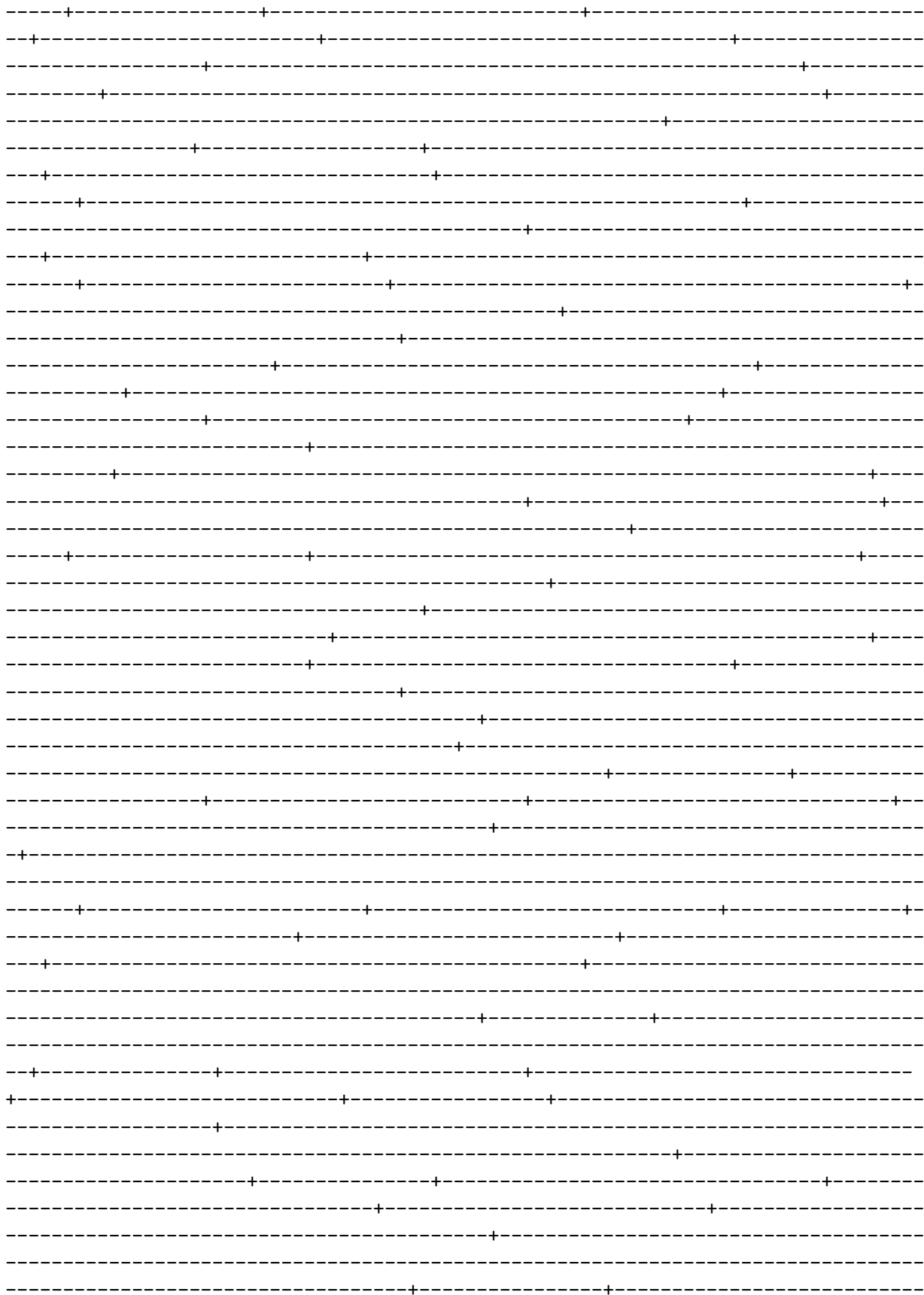


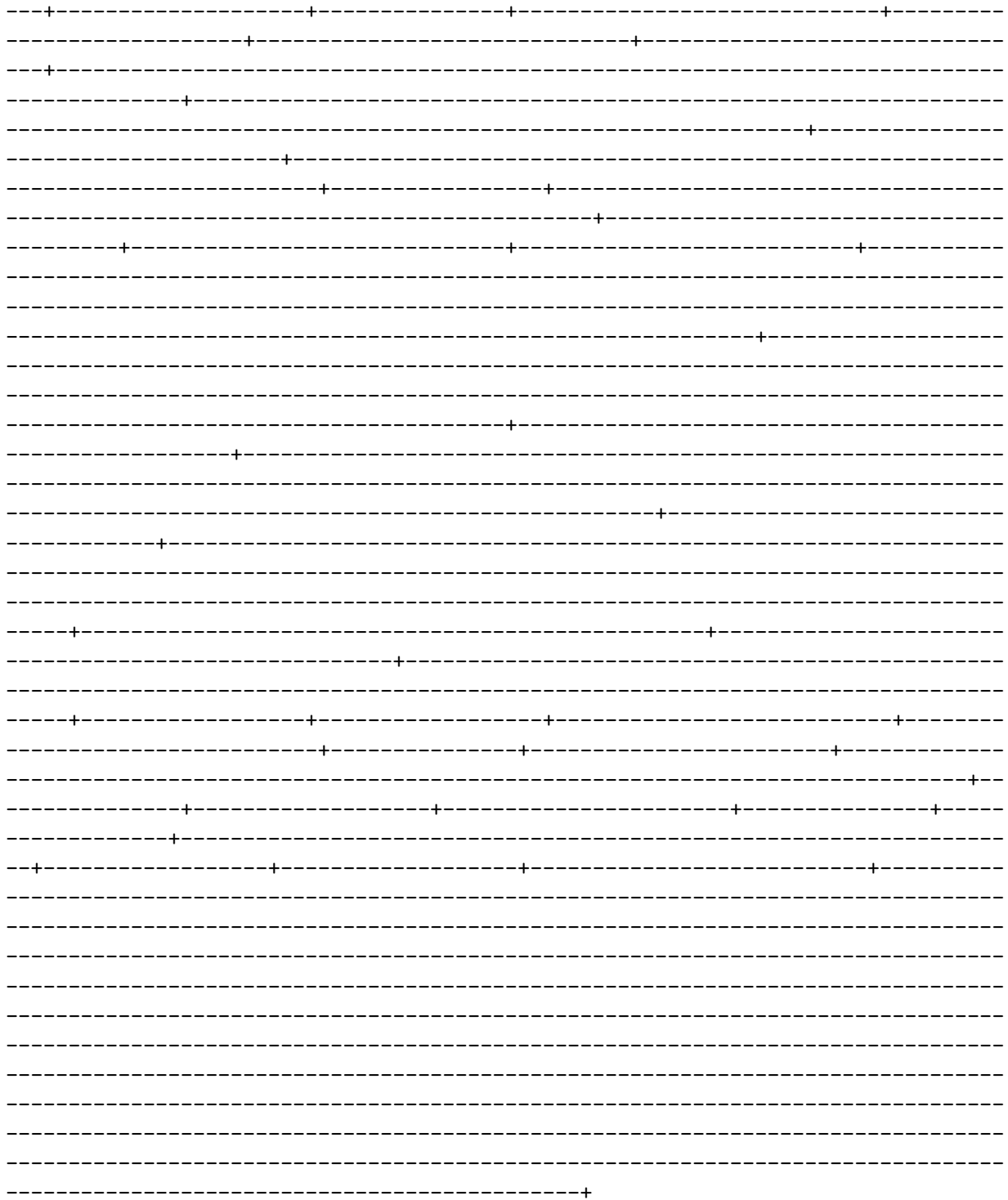












	count	
0		
0	0	0 0 0 0 0 0
0		0 0
0		0
0	0	
0		0 0 0 0
0	0	

The image displays a 50x50 grid of 2500 '0' characters, each followed by a vertical bar. The characters are arranged in a sparse, non-uniform pattern across the grid. The distribution is not random; it appears to follow a specific, though complex, rule or algorithm. The characters are more densely packed in certain areas, such as the top-left and bottom-right, and more sparse in others. The overall effect is a textured, almost abstract representation of a binary or categorical data set.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

null				null
null	null			
null		null		
null			null	
null				null
null		null		null
null				
null				
null				
null	null		null	null
null			null	
null		null		null
null		null	null	
null	null			
null	null		null	
null		null	null	
null				
null			null	null
null		null		
null		null	null	null
null	null		null	null
null		null		null
null		null		
null			null	
null				
null	null		null	null
null		null		null
null	null	null		null
null		null		null
null		null	null	
null		null	null	
null			null	null
null		null		null
null	null			
null		null		
null		null	null	
null			null	null
null				
null	null			null
null			null	
null		null		
null		null		null
null			null	
null				
null		null		
null	null			null
null		null		
null				

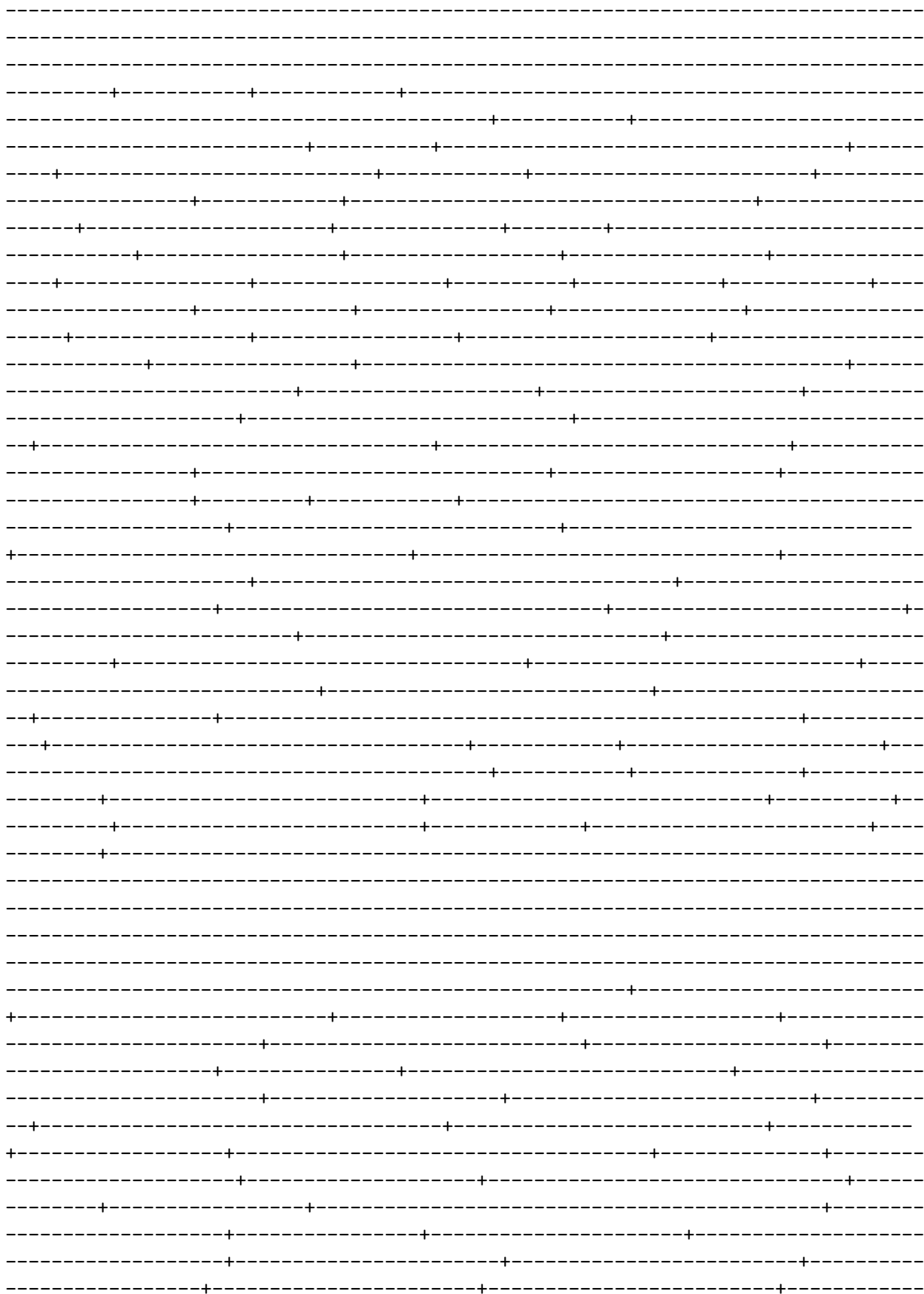
null		null		null
null		null		
null			null	
null				null
null		null		
null		null	null	
null			null	
null				null
null		null		null
null				
null				
null	null		null	null
null			null	
null		null		null
null		null	null	null
null				
null	null			
null	null		null	
null		null	null	
null				
null		null	null	
null		null		null
null				
null	null		null	null
null		null	null	
null		null		
null				
null			null	
null	null			null
null		null	null	
null		null		null
null		null	null	null
null		null		
null			null	null
null		null		null
null				
null			null	
null	null		null	
null		null		null
null			null	null
null				
null				
null	null			null
null		null	null	
null				
null			null	null
null				
null	null			null
null			null	
null				
null			null	

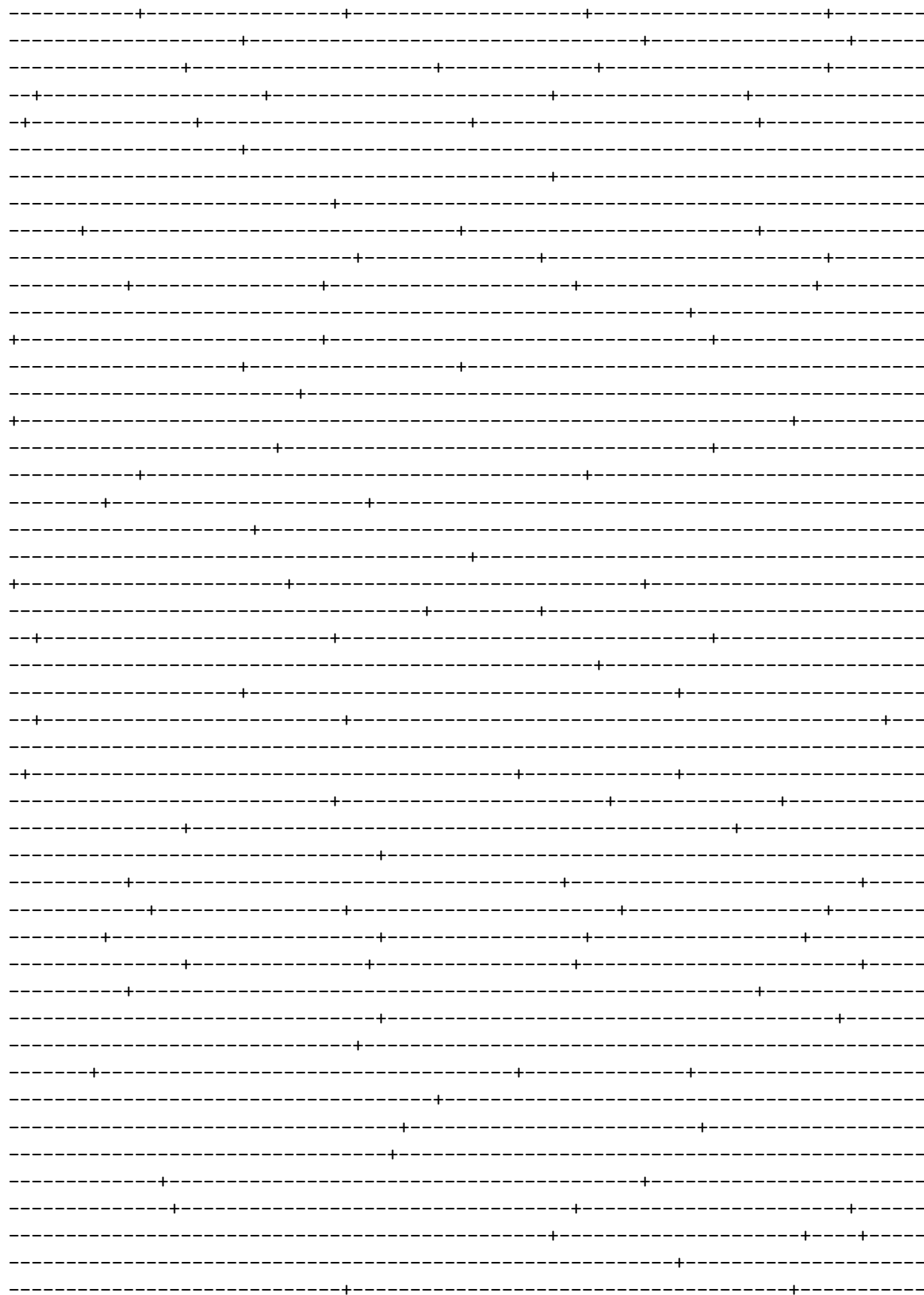
null		null		null	
null		null			null
null				null	
null			null		null
null		null		null	
null				null	
null					null
null	null				
null		null			
null			null		
null					null
null		null			null
null					
null	null		null		null
null			null		
null		null		null	null
null		null		null	
null	null				
null	null		null		
null		null	null		
null			null	null	
null			null		null
null	null		null		null
null		null		null	
null		null			
null				null	
null	null		null		null
null		null			null
null	null		null		null
null		null		null	null
null		null		null	
null			null		null
null	null		null		null
null		null		null	
null			null		null
null				null	

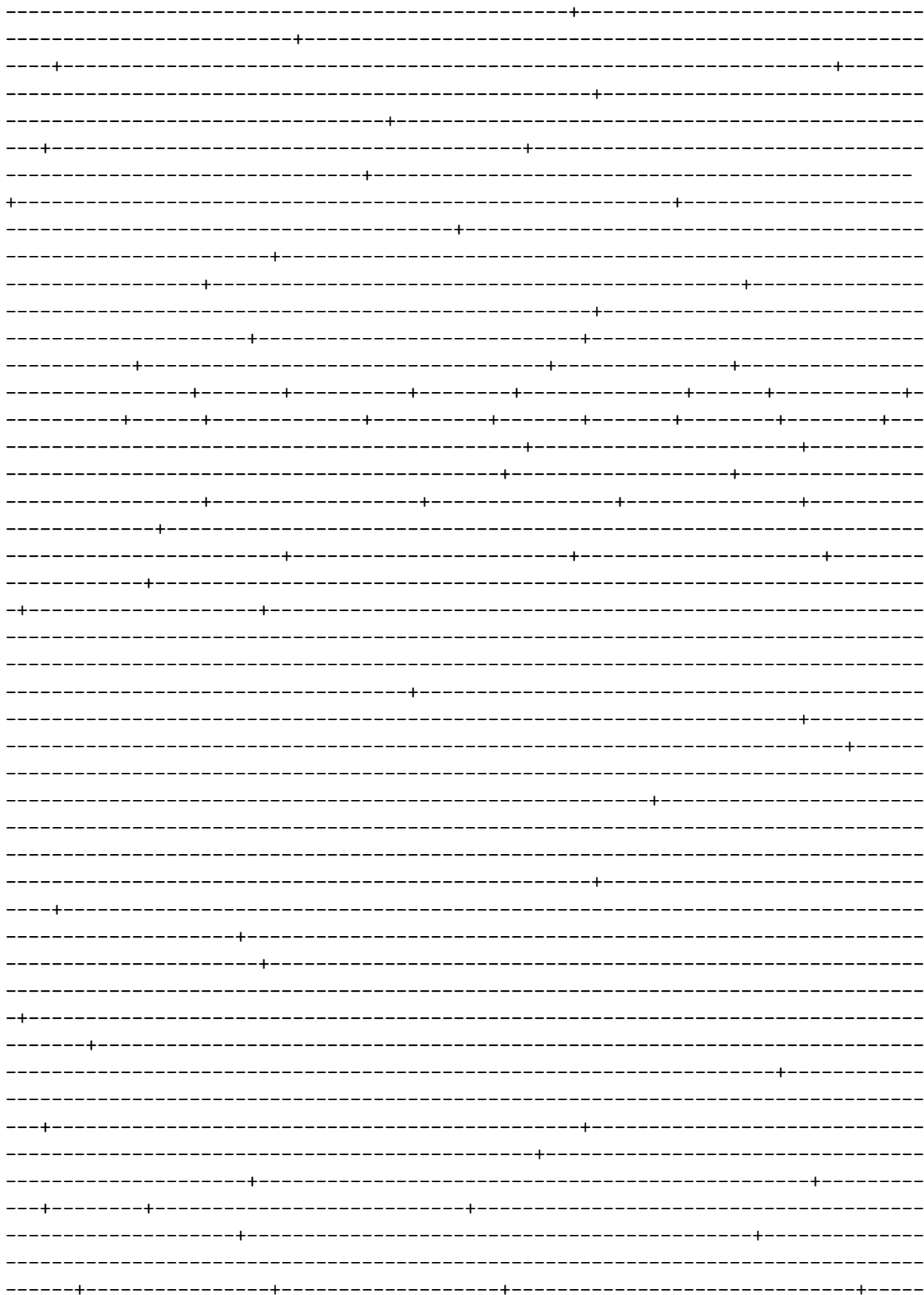
null		null			null
null				null	
null					null
null		null			
null		null			null
null					null
null			null		null
null			null		
null				null	
null					null
null		null			
null		null		null	
null				null	
null					null
null		null			null
null					
null	null		null		null
null			null		
null				null	null
null		null		null	
null		null			
null	null		null		
null	null		null		
null		null		null	
null				null	null
null			null		null
null			null		
null	null		null		null
null		null		null	
null		null			
null				null	
null					null
null	null		null		null
null		null		null	
null				null	null
null			null		null
null					
null				null	
null	null		null		null
null		null		null	
null		null			null
null			null		null
null				null	
null					null
null	null		null		
null				null	
null					null
null	null		null		

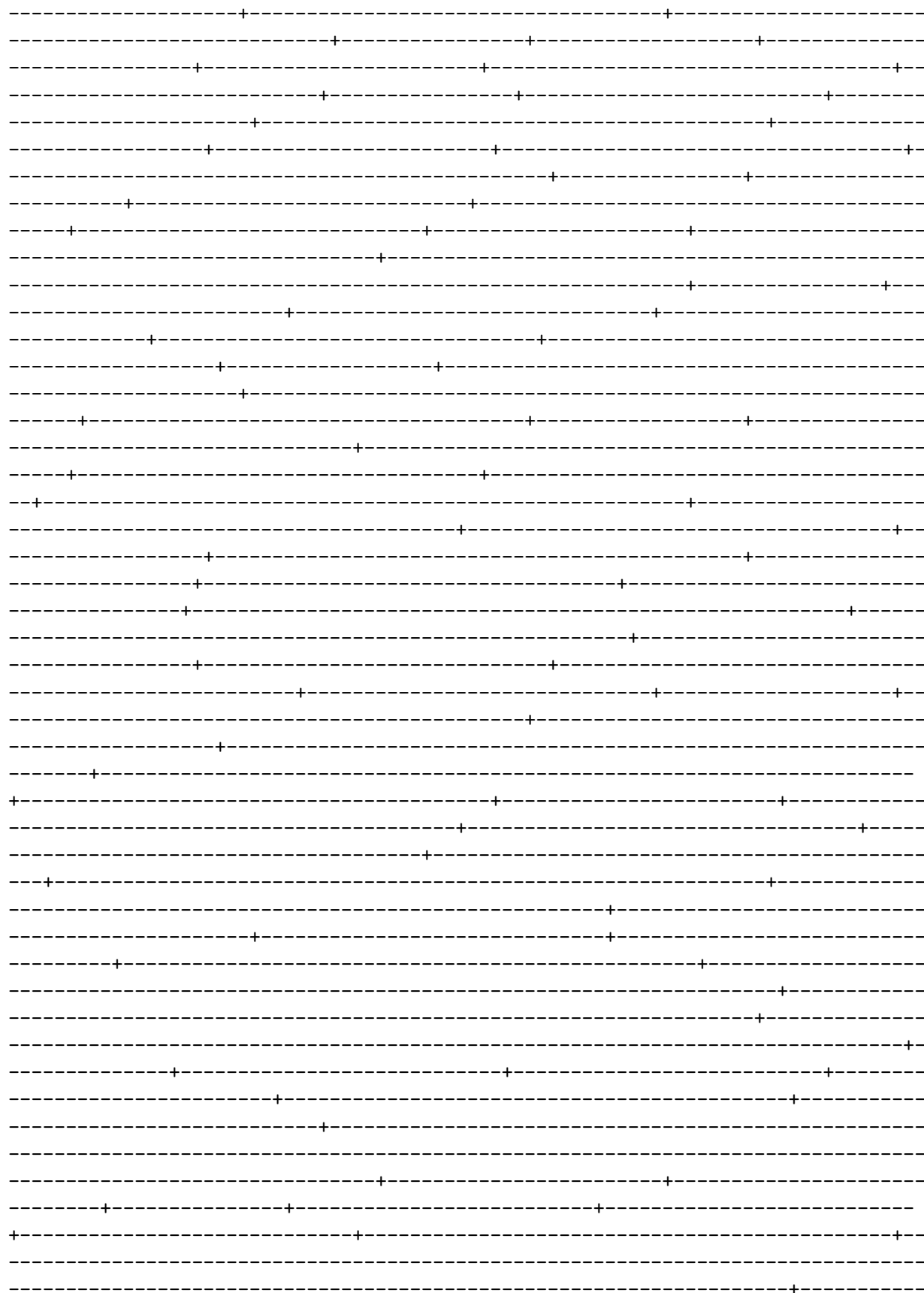
null		null		null
null			null	null
null				null
null	null			null
null			null	
null				null
null		null		
null	null			null
null			null	
null		null		null
null			null	
null				null
null		null		
null		null	null	
null			null	
null				null
null		null		
null			null	
null		null		null
null				null
null	null		null	null
null			null	
null		null		null
null		null	null	
null				null
null			null	
null		null		null
null			null	
null		null		null
null			null	
null	null		null	null
null		null		null
null		null		
null			null	
null				null
null		null		
null	null			
null			null	
null				null
null			null	
null				null
null			null	
null				null
null			null	
null				null
null	null	null		null

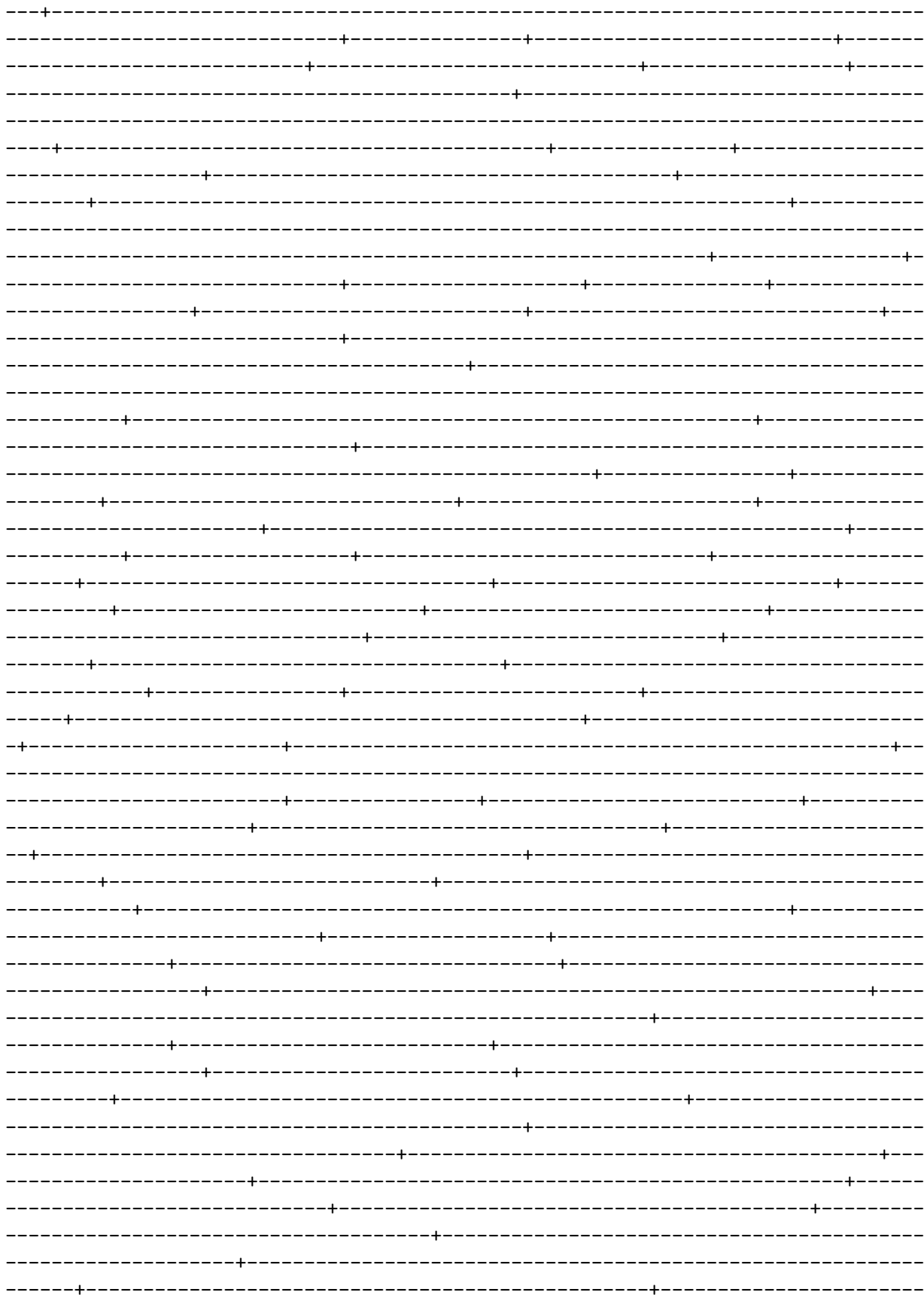
null	null	null	
null	null	null	null
null	null		
null	null	null	null
null			

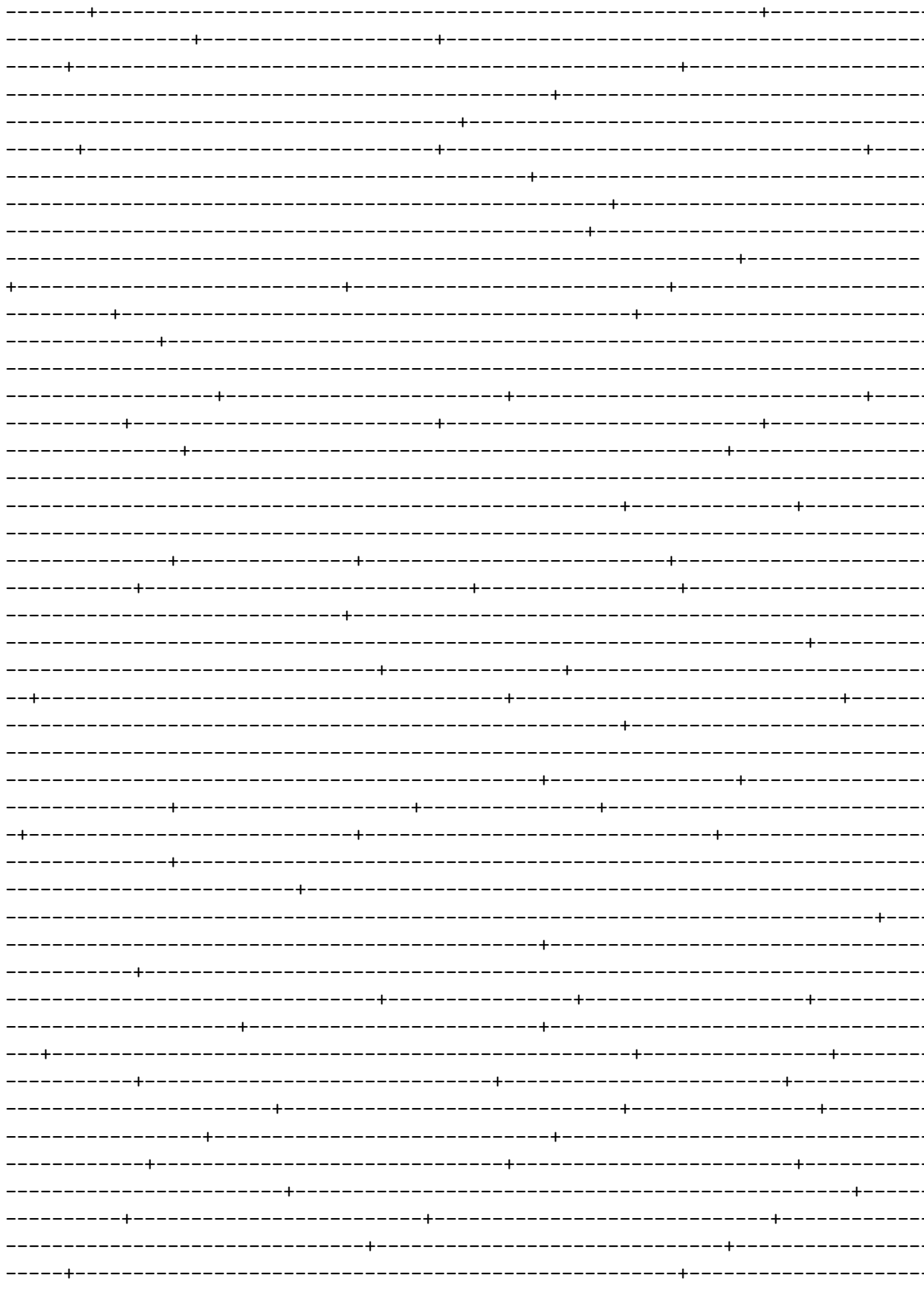


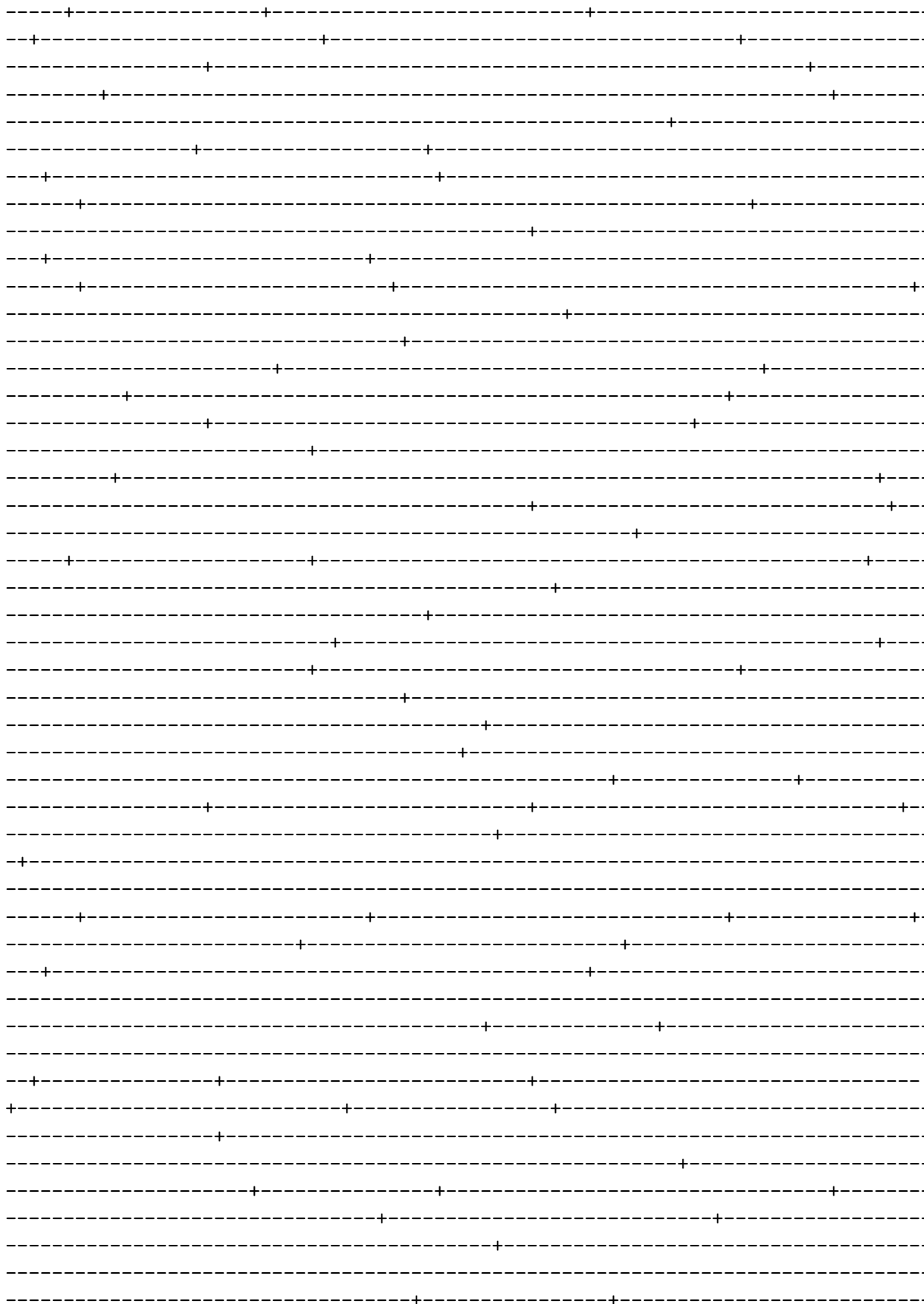


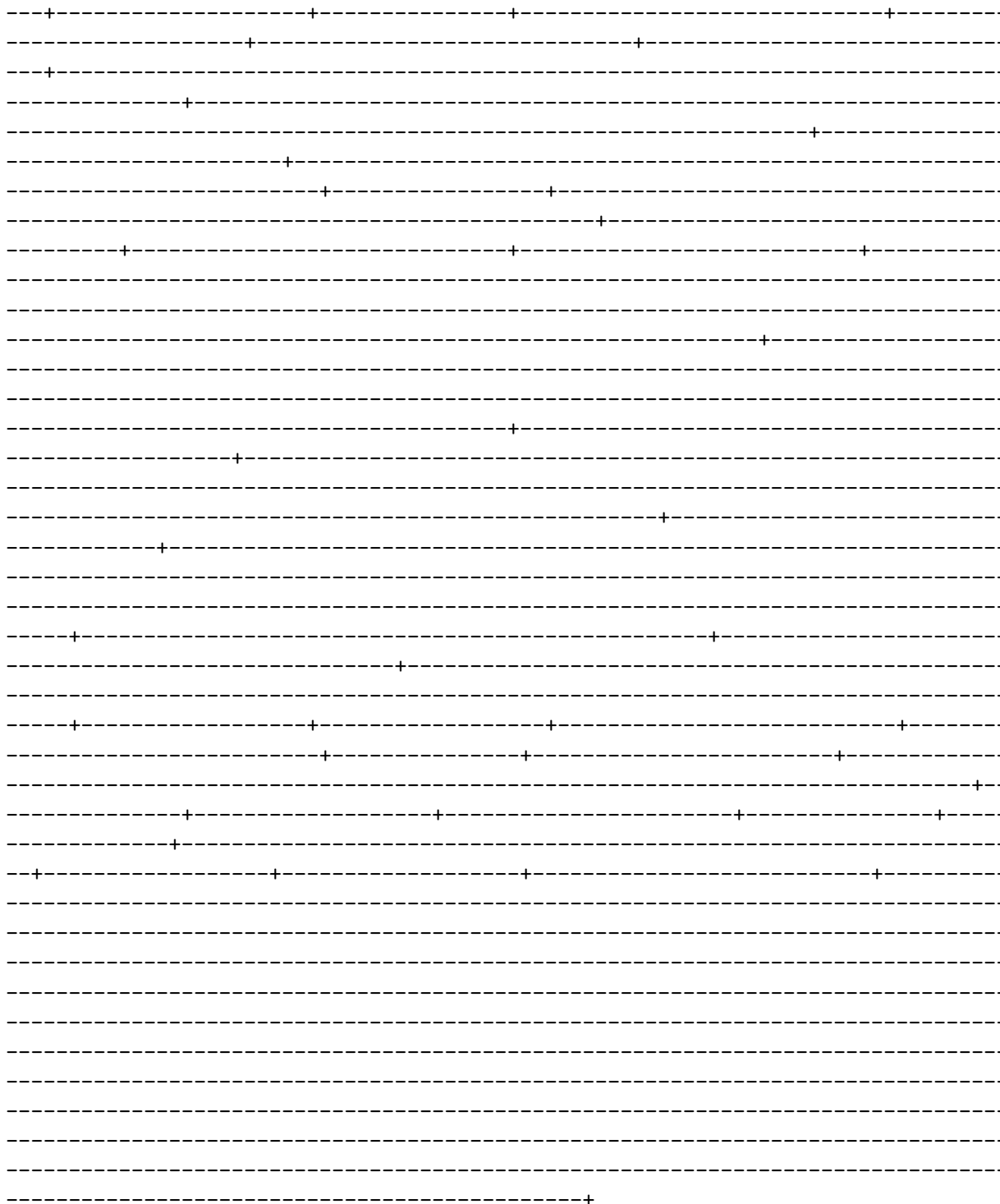












[]:

0.4 4. Hacerle cambios al Data Frame en Spark

```
[ ]: Quitando los duplicados
```

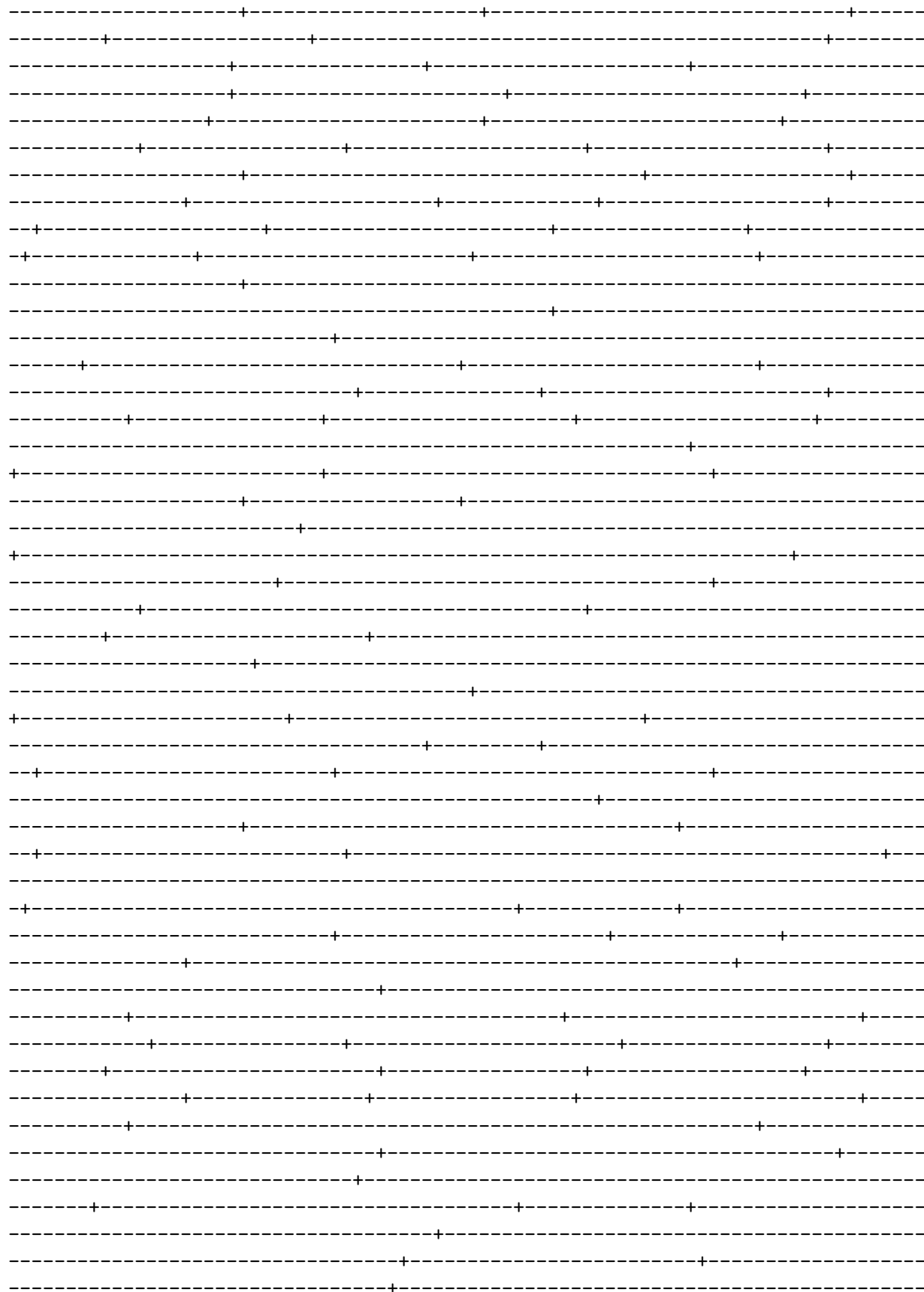
```
[36]: sub_nd_british2_df = british2_df.dropDuplicates()
```

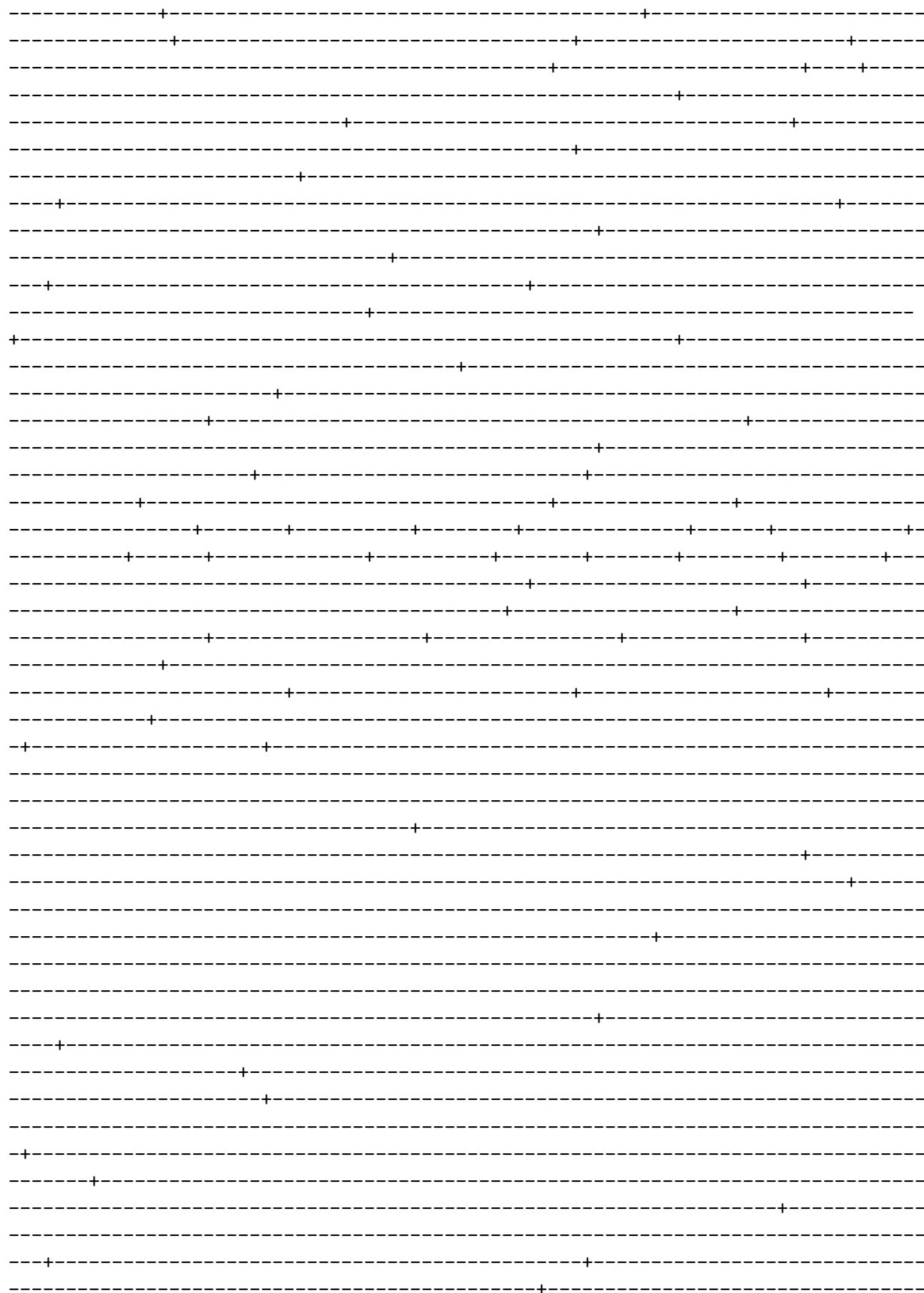
```
[37]: sub_nd_british2_df.describe().show()
```

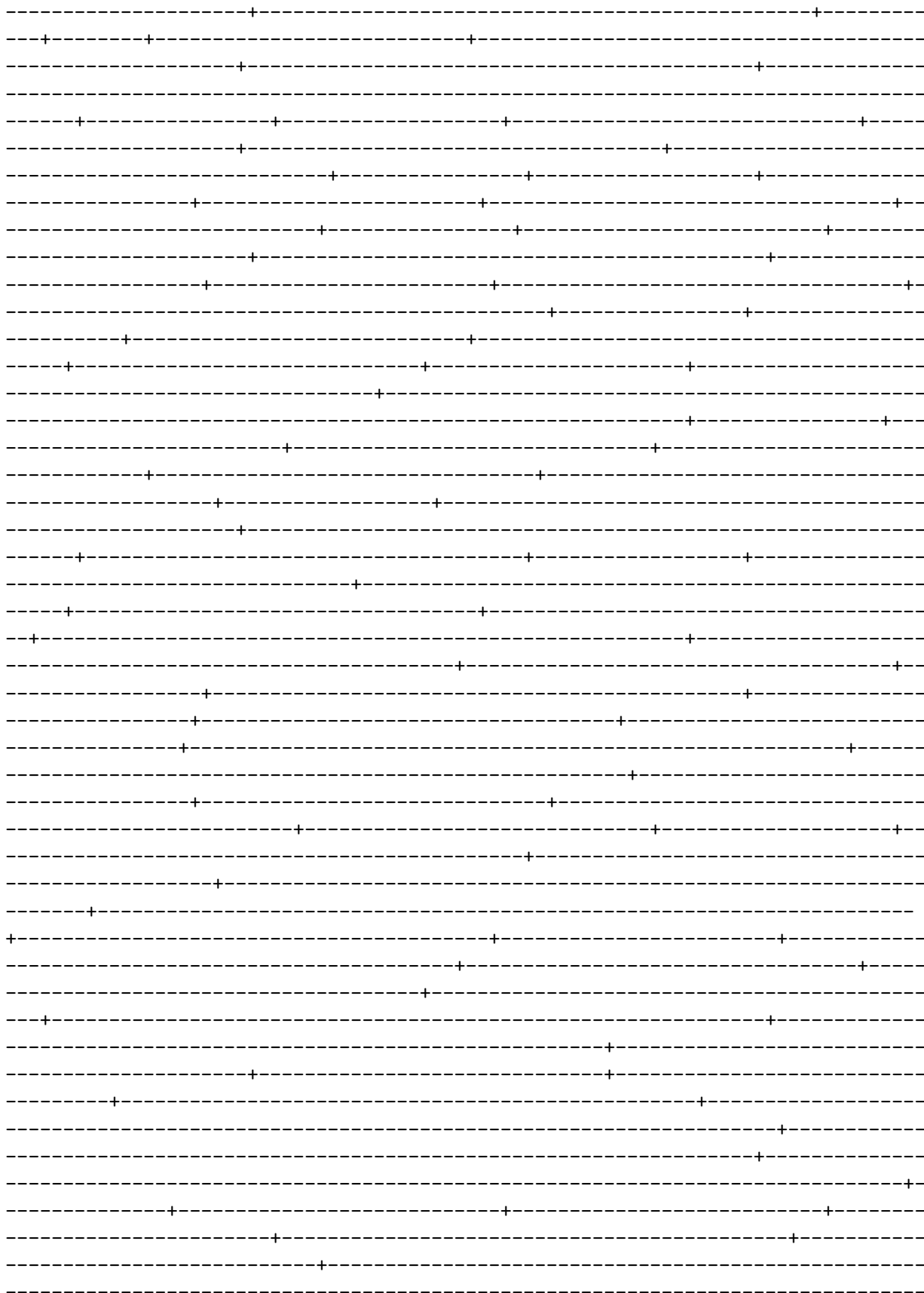
```
23/09/22 08:47:33 WARN DAGScheduler: Broadcasting large task binary with size
5.8 MiB
```

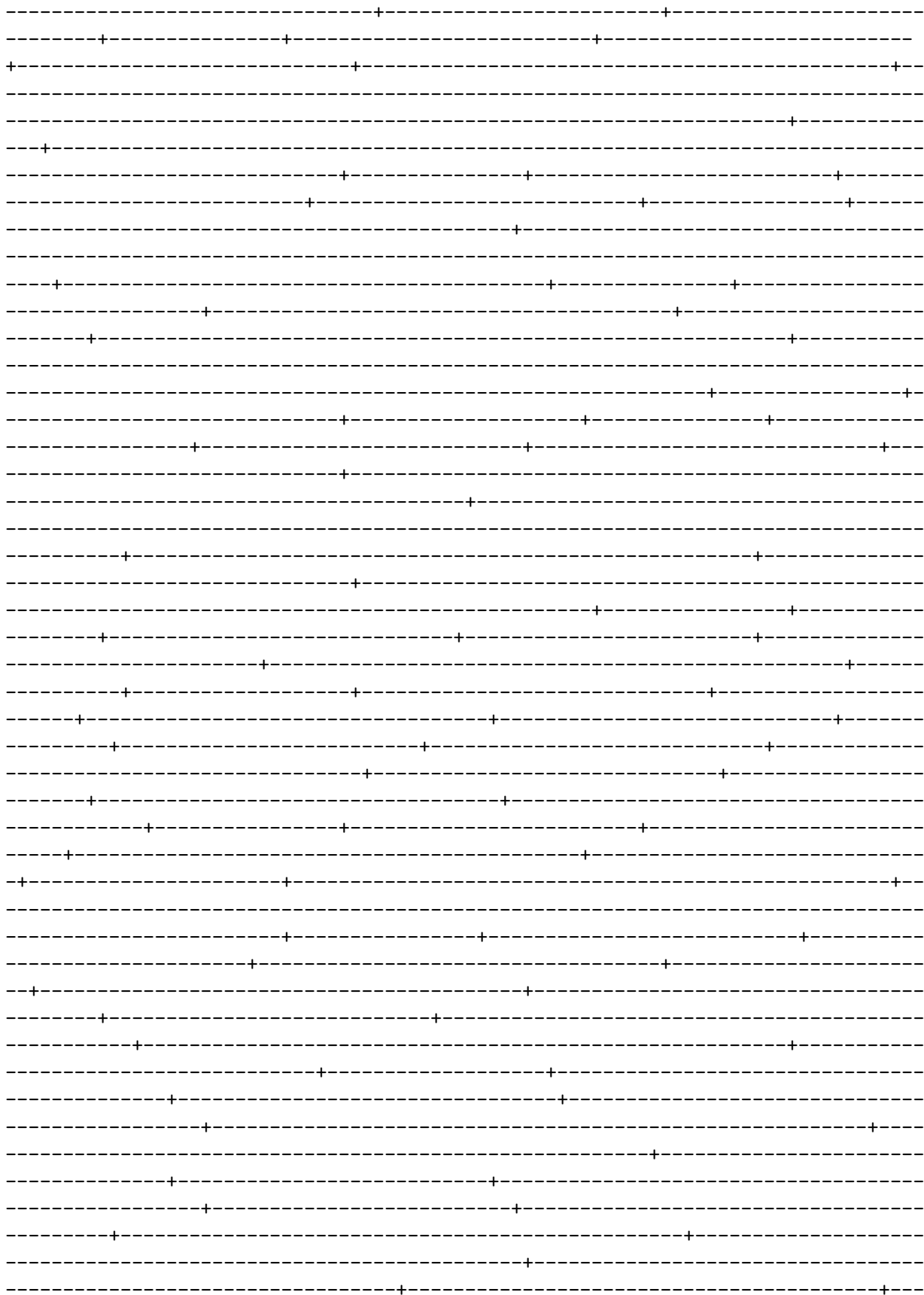
[illegible]

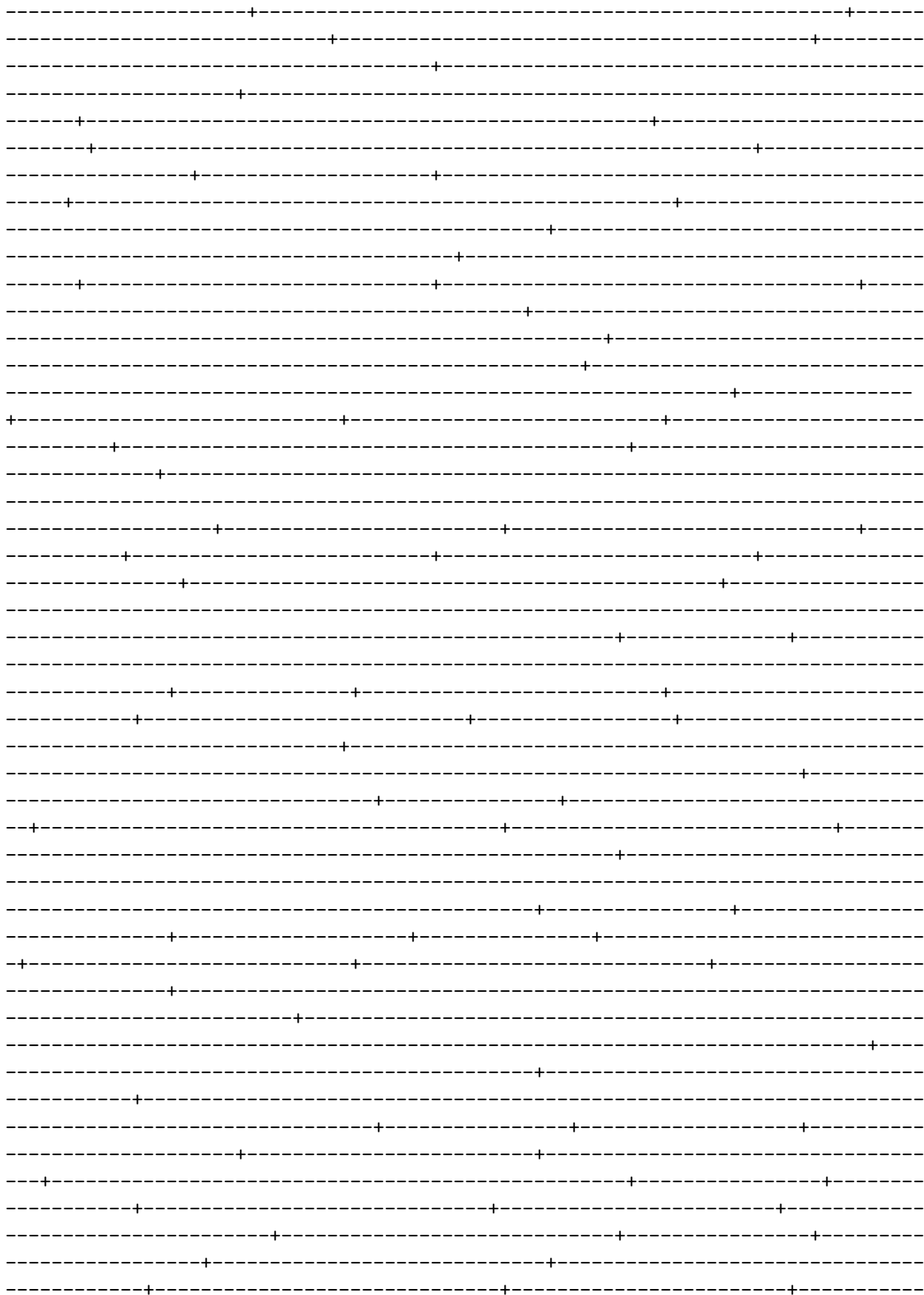


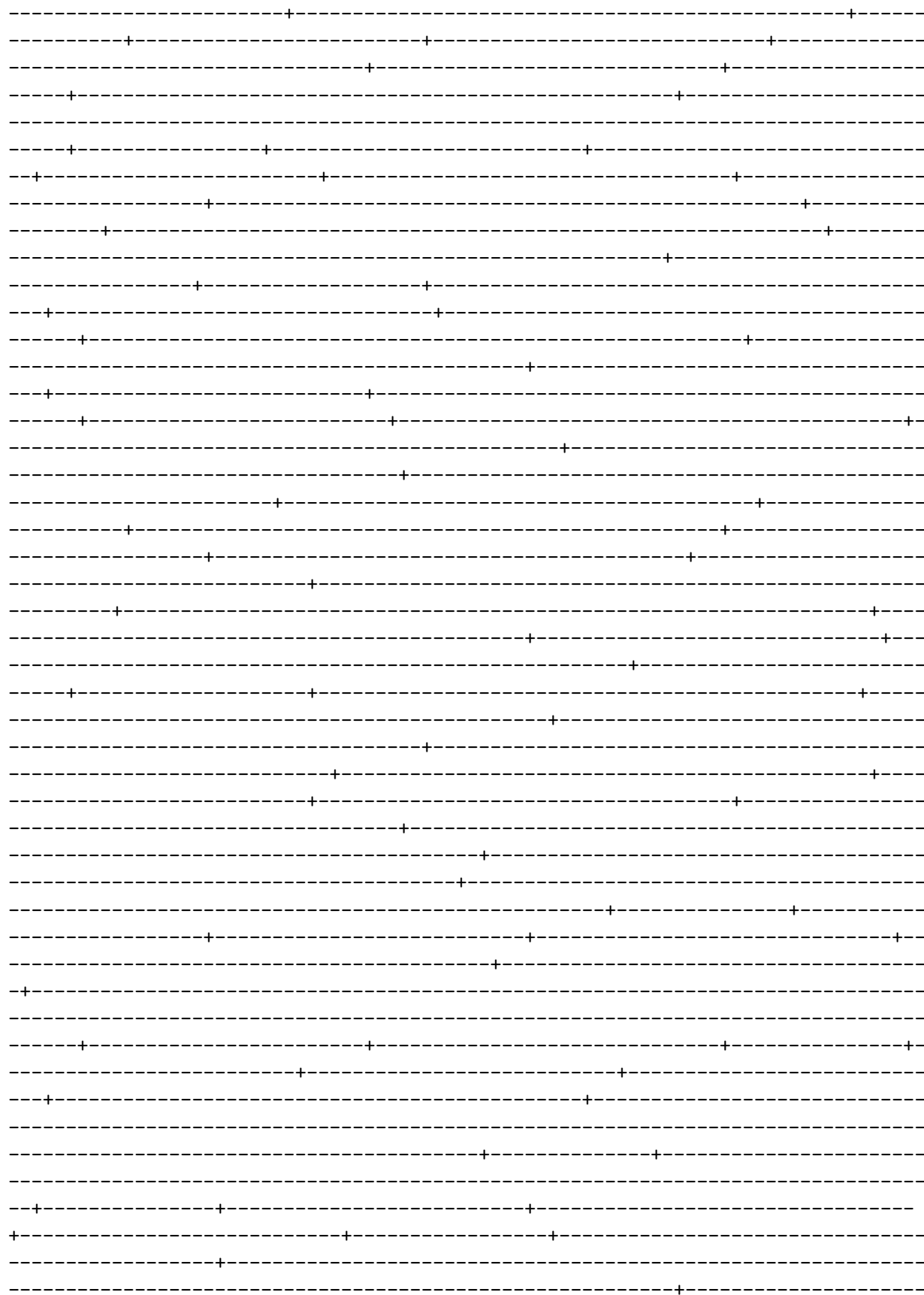













```

content="width=device-width|initial-scale=1"><meta name="referrer"
content="origin-when-cross-origin"><meta name="og:image"
content="https://d2ogkq1rg66kh0.cloudfront.net/site-resources/images/fb-image-
share.3a87a6eb.png"><meta name="og:site_name" content="data.world"><meta
name="twitter:card" content="summary"><meta name="twitter:site"
content="@datadotworld"><meta property="fb:app_id"
content="776416075825233"><link rel="stylesheet" href="https://fonts.googleapis.
com/css2?family=Inconsolata:wght@400;700&family=Lato:ital|wght@0|300;0|400;0|700
;1|300;1|400;1|700&display=swap"><link rel="search" href="/opensearch.xml"
type="application/opensearchdescription+xml" title="data.world"> <link
rel="apple-touch-icon" sizes="57x57"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-57x57.png"><link rel="apple-touch-icon" sizes="60x60"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-60x60.png"><link rel="apple-touch-icon" sizes="72x72"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-72x72.png"><link rel="apple-touch-icon" sizes="76x76"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-76x76.png"><link rel="apple-touch-icon" sizes="114x114"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-114x114.png"><link rel="apple-touch-icon" sizes="120x120"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-120x120.png"><link rel="apple-touch-icon" sizes="144x144"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-144x144.png"><link rel="apple-touch-icon" sizes="152x152"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-152x152.png"><link rel="apple-touch-icon" sizes="180x180"
href="https://assets.data.world/icons-d1940ad25da01d05b8f27389101617c1/apple-
touch-icon-180x180.png"><meta name="apple-mobile-web-app-capable"
content="yes"><meta name="apple-mobile-web-app-status-bar-style" content="black-
translucent"><meta name="apple-mobile-web-app-title" content="data.world"><meta
name="mobile-web-app-capable" content="yes"><meta name="theme-color"
content="#fff"><meta name="application-name" content="data.world"><link
rel="icon" type="image/png" sizes="32x32" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-32x32.png"><link rel="icon"
type="image/png" sizes="16x16" href="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon-16x16.png"><link rel="shortcut icon" hr
ef="https://assets.data.world/icons-
d1940ad25da01d05b8f27389101617c1/favicon.ico"><link
href="https://assets.data.world/css/vendors~app.bf2ab82ca4e8e237d22a.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors~vendors.e418760f21d557e2e0e2.css"
rel="stylesheet"><link
href="https://assets.data.world/css/vendors.5b2975e4af1441b468c5.css"
rel="stylesheet"><link
href="https://assets.data.world/css/app.1915b3c34db820707a63.css"
rel="stylesheet"></head><body ><div id="root"><div class="page-container
CoreLayout__embed__1fxWq"><div class="view-container"><div

```



```

1}|"siteStatus":{"uploadsOperational":true|"queryOperational":true|"websiteOperational":true|"downloadsOperational":true|"shouldAskReload":false|"shouldForceReload":false|"connectionError":false|"incidentName":"","incidentMessage":"","pageStatus":"none"}|"treatments":{"isSite":false|"activeTreatments":["showVersions"|showWorkspaceTabs|showEntityLinkOptions|canCreateOpenDatasets|canCreateOrgs|showConversionPages|showSignup|"showSocialLogin"]}|"user":{"hasError":false|"loadedUsers":{"userFollows":{"userLikes":{"userRecentComments":{"groupAuthorizations":{"featuredDatasets":{"loaded":false|"resources":[]}}}}</script><script type="application/json" id="init-config">{"adwords_pixel_id":"869213640|"analytics":{"segment_id":"IXhtZmPohuGM60VHk59cdSbWqWQBB7aR|"segment_endpoint":"https:\u002F\u002Fapi-el.data.world"}|"analytics_log_to_console":false|"base_cookie_domain":null|"contact_admin_form_config":[{"type":"input|"label":"Email|"lookupProperty":"email|"readOnly":true162|"required":true}|{"type":"textarea"164|"label":"Message"165|"placeholder":"Add a message (i.e details that need to be reviewed)"|"required":true}]167|"csrf_cookie_name":"_csrf|"csrf_header_name":"x-csrf-token|"data_provider_agents":["johnsnowlabs|"associatedpress"]|"download_content_url":"https:\u002F\u002Fdownload.dwcontent.com|"download_url":"https:\u002F\u002Fdownload.data.world|"dw_tableau_wdc_host_domain":"https:\u002F\u002Ftableau.data.world|"deployed_environment_full":"prod|"facebook_pixel_id":"261984994177057|"fb_app_id":"776416075825233|"google_chouser_app_id":"956984791354|"help_cards":[{"id":"getting_started|"header":"Get Started|"body":"New to data.world? Start here to find introductory articles."|"link":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html|"imageSrc":"getting-started-lightglow"}|{"id":"join_discussion|"header":"Join the discussion|"body":"Join the general discussion or start your own topic|"link":"","imageSrc":"discussions-lightglow"}|{"id":"see_documentation|"header":"Visit the Help Center|"body":"Find general platform documentation and technical tutorials."|"link":"https:\u002F\u002Fdocs.data.world\u002F"|"imageSrc":"documentation-lightglow"}|{"id":"contact_support|"header":"Contact support|"body":"Report an issue or make a suggestion."|"link":"https:\u002F\u002Fhelp.data.world\u002Fhc\u002Fen-us\u002Frequests\u002Fnew"|"imageSrc":"support-lightglow_sales-guy"}]|"hoot_url":"","iframely_key":"e4239223dd4ad21531b3a044840d5898|"is_monitored_env":true|"is_vpc":false|"max_upload_size":524288000|"media_upload_domain":"mediauploads.data.world|"media_url":"https:\u002F\u002Fmedia.data.world|"oauth_authorize_url":"https:\u002F\u002Fauth.data.world|"oauth_providers":[{"name":"facebook|"display_name":"Facebook|"active":true}209|{"name":"github|"display_name":"GitHub|"active":true}212|{"name":"google|"display_name":"Google|"active":true}215|{"name":"linkedin|"display_name":"LinkedIn|"active":false}|{"name":"twitter|"display_name":"Twitter|"active":true}]|"openAccessRoutePrefixes":["\u002Fplaineale rprojects"|"\u002Fd1gi\u002Fmissing-fb-posts-w-share-stats"|"\u002Ftravelwarnings\u002Ftravel-danger"|"\u002Ffootprint\u002Fnfa-2018-edition"|"\u002Fpolymathic\u002Fcasualties-of-the-syrian-civil-war"|"\u002Faurielle\u002Ffortune-500-2017"|"\u002Fnjdotcom"|"\u002Fian\u002Fusa-baby-names-national-and-state-by-state"|"\u002Fowentemple\u002Ftext-and-content-

```


features-of-most-persuasive-ted-talks|\u002Fschnee\u002Ftangled|\u002Fchhs\u002F
 fc544658-35c5-4be0-af20-fc703bc57c13|\u002Fnz-stats-
 nz\u002F4e5ebed0-4279-4292-a85f-33c9f2a93f8d|\u002Fshad\u002Fcovid-19-time-
 series-data|\u002Fmarkmarkoh\u002Fcoronavirus-data|\u002Fdallemang\u002Fcovid-
 19-data|\u002Fpolicinginamerica|\u002Findustryresearch\u002Fdata-engineer-
 survey-2021|\u002Fembed\u002Fswwo|swwo|\u002Fassociatedpress\u002F2018-election-
 results-precinct-level|\u002Faaron-kessler-ap\u002Fcongressional-
 earmarks|\u002Fassociatedpress\u002Fap-
 votecast|\u002Fassociatedpress\u002Fbolsoa-familia-aid-to-brazils-
 poor|\u002Fassociatedpress\u002Fcdc-covid-19-vaccine-
 tracker|\u002Fassociatedpress\u002Fcovid-impact-survey-public-
 data|\u002Fassociatedpress\u002Fcovid-tracking-project-testing-in-
 states|\u002Fassociatedpress\u002Fap-planner-covid-reopening-
 data|\u002Fassociatedpress\u002Ffema-distribution-of-ppe-to-
 states|\u002Fassociatedpress\u002Fforeign-exchange-
 rates|\u002Fassociatedpress\u002Fus-dams-
 public|\u002Fassociatedpress\u002Fjohns-hopkins-coronavirus-case-
 tracker|\u002Fassociatedpress\u002Fmass-killings-
 public|\u002Fassociatedpress\u002Fopioid-
 prescriptions-2010-2015|\u002Fassociatedpress\u002Fpublic-health-official-
 departures|\u002Fassociatedpress\u002Frefugee-admissions-to-us-end-
 fy-2018|\u002Fassociatedpress\u002Frental-assistance-from-coronavirus-relief-
 fund|\u002Fassociatedpress\u002Fstate-ppe-
 purchases|\u002Fassociatedpress\u002Fmarshall-project-covid-cases-in-
 prisons|\u002Fassociatedpress\u002Fvehicle-miles-traveled|\u002Ffrare\u002Ffish-
 surveys|" \u002Ffrare\u002Fflarval-dispersal-data"]|"preferred_oauth_provider":"goo
 gle"|"prod_env":true|"profile_interests":["education"|finance|government|busines
 s|transportation|health|agriculture|environment|sports|entertainment|technology|
 science|culture|politics|"crime"]|"query_url":"https:\u002F\u002Fquery.data.world"
 "regional_cluster":"US"|"request_dataset_access_form_config":[{"type":"input"
 "label":"Full name"|"lookupProperty":"displayName"284|"required":true285|"readO
 nly":true}]{ "type":"email"|"label":"Contact email"|"placeholder":"Provide an
 email in the case a dataset administrator needs to reach out"289|"showOptionalFl
 ag":true}]{ "type":"textarea"291|"label":"Message"292|"placeholder":"Share what
 you would like to do with this dataset"293|"required":true}]294|"requests":{"con
 tributionSuggestionInfoText":"You are suggesting changes to this \u003C%=
 resourceType %\u003E that will need to be approved by an admin of this
 organization. Only you and admins will be able to see your changes until they're
 approved."|"pendingApprovalsText":"There \u003C%= pendingApprovalsNumber %\u003E
 from the community pending admin approval."|"pendingSuggestionsText":"Your
 suggested changes are visible on the Suggestions tab| but only you and this
 organization's admins can see them until they have been approved. The Overview
 tab displays only approved information."|"personalResourceSharedWithText":"This
 draft \u003C%= resourceType %\u003E is within your private collection and only
 visible to you. It is neither discoverable nor viewable by the data.world
 community. To open it up to others to view| publish it to an
 organization."|"transferOwnershipMessage":"By publishing this \u003C%=
 resourceType %\u003E to an organization| you are transferring ownership to the

organization. Once your draft is published| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."|"transferOwnershipModalWarningBanner":"Once the organization has published your draft| you will no longer be able to directly edit this \u003C%= resourceType %\u003E unless you become a content creator of that organization."|"sentry_dsn_client":"https:\u002F\u002Ff763f0cf14f04c75b4789a118ab1beb3@sentry.io\u002F1486857"|"server_origin":"https:\u002F\u002Fdata.world"|"sparql_host":"https:\u002F\u002Fpdw-sparqlproxy-us-east-1.prod.data.world"|"stripe_public_key":"pk_live_CPGU5gVFxlkY6pabDXladsMR"|"swagger_use_case_label_override_org":["oa-test"|engie-impact|"e-ddw"]|"tableau_api_version":"2.8"|"ui":{"orgAdmin":{"resourceLinks":[{"key":"getting-started"|"linkDisplayName":"Getting started guide"315|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"316|"newTab":true317|"svgIcon":"meta"318|"verticallyAlignIcon":true}319|{"key":"documentation"320|"linkDisplayName":"Documentation"321|"url":"https:\u002F\u002Fdocs.data.world\u002F"322|"newTab":true323|"svgIcon":"meta"324|"verticallyAlignIcon":true}325|{"key":"integrations"326|"linkDisplayName":"Integrations"327|"url":"\u002Fintegrations"328|"newTab":true329|"svgIcon":"integration"330|"verticallyAlignIcon":true331|"showIfAnyDismissed":["onboard-checklist"332|"onboard-checklist-admin"]}}333|{"key":"new-features"334|"linkDisplayName":"New features"335|"url":"https:\u002F\u002Fwhatsnew.data.world"336|"newTab":true337|"svgIcon":"magic-wand"338|"verticallyAlignIcon":true339|"showIfAnyDismissed":["onboard-checklist"340|"onboard-checklist-admin"]}}341|{"key":"enterprise"342|"linkDisplayName":"data.world for the enterprise"343|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-demo\u002Fsrc=inapp_data.world-for-the-enterprise"344|"newTab":true345|"svgIcon":"organization"346|"verticallyAlignIcon":true}}347|"onboardChecklist":{"key":"site"348|"items":[{"id":"createAccount"349|"analyticsId":"onboard_checklist.create_account"350|"complete":true351|"formattedMessage":{"id":"onboard_checklist.create_account"352|"description":"Text for onboard checklist create account link"353|"defaultMessage":"Create account"}}354|{"id":"inviteTeam"|"analyticsId":"onboard_checklist.invite-team"|"href":"\u002F\u003C%- orgAgentid %\u003E?tab=members"|"dismissOnClick":"invite-team-link"|"completeIfDismissed":"invite-team-link"|"formattedMessage":{"id":"onboard_checklist.invite-team"|"description":"Text for onboard checklist invite team link"|"defaultMessage":"Invite your team"}}|{"id":"connectData"|"analyticsId":"onboard_checklist.connect-data"|"href":"\u002Fcreate-a-dataset"|"dismissOnClick":"connect-data-link"|"completeIfDismissed":"connect-data-link"|"formattedMessage":{"id":"onboard_checklist.connect-data"|"description":"Text for onboard checklist connect data link"|"defaultMessage":"Connect data sources"}}|{"id":"exploreIntegrations"371|"analyticsId":"onboard_checklist.explore-integrations"372|"href":"\u002Fintegrations"373|"target":"_blank"374|"dismissOnClick":"explore-integrations-link"375|"completeIfDismissed":"explore-integrations-

```

link"376|"formattedMessage":{"id":"onboard_checklist.explore-
integrations"377|"description":"Text for onboard checklist explore integrations
link"378|"defaultMessage":"Explore integrations"}}|{"id":"createProject"380|"ana
lyticsId":"onboard_checklist.create-project"381|"href":"\u002Fcreate-a-
project"382|"dismissOnClick":"create-project-
link"383|"completeIfDismissed":"create-project-
link"384|"formattedMessage":{"id":"onboard_checklist.create-
project"385|"description":"Text for onboard checklist create project
link"386|"defaultMessage":"Create your first
project"}}}|{"onboardTour":[{"show":true388|"header":"Welcome to
data.world!"389|"subhead":"The place for modern data
teamwork."390|"leftCol":"Organize data work so you and your team find what you
need&nbsp;faster."391|"middleCol":"Explore and create a shared understanding of
your&nbsp;data."392|"rightCol":"Share insights and invite more people to
participate in data&nbsp;teamwork."}393|{"show":true394|"imageHeaderKey":"search
"395|"imageKey":"collaborate"396|"imageAlt":"Start with Data"397|"header":"Start
with fantastic&nbsp;data"398|"subhead":"Browse thousands of free399| open data
resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from
almost anywhere into a single searchable&nbsp;library.\n\nMake it open400| share
with your team401| or keep it&nbsp;private."}402|{"show":true403|"imageHeaderKey
":"bulb"404|"imageKey":"dataProject"405|"imageAlt":"Data
Projects"406|"header":"Organize your data&nbsp;projects"407|"subhead":"Create
projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd
context to make it easier for others to discover408| understand409| and
contribute to your&nbsp;projects.\n\nImmediately query data without leaving your
&nbsp;projects."}410|{"show":true411|"imageHeaderKey":"code"412|"imageKey":"slid
eIntegrations"413|"imageAlt":"Integrations"414|"reverse":true415|"header":"Conne
ct the data tools you&nbsp;love"416|"subhead":"Automatically keep data.world
synced with your data&nbsp;sources.\n\nIntegrate your team's favorite tools417|
from analysis to&nbsp;visualization."}418|{"show":true419|"imageHeaderKey":"disc
ussions"420|"imageKey":"discussionsCollaboration"421|"imageAlt":"Contributors"42
2|"header":"Add more brainpower to&nbsp;your&nbsp;projects"423|"subhead":"Invite
your team to securely and effectively work&nbsp;together.\n\nManage fine-grained
permissions for each&nbsp;member.\n\nAsk questions424| add insights425| bookmark
top contributions426| and&nbsp;more."}427|{"show":true428|"imageHeaderKey":"insi
ght"429|"imageKey":"knowYourData"430|"imageAlt":"know your
data"431|"header":"Share your results"432|"subhead":"Share meaningful insights
and the data behind&nbsp;them.\n\nDistribute clean433| documented data for
others to use in their own&nbsp;work.\n\nHelp the right people discover the
resources they&nbsp;need."}]}434|"orgMember":{"resourceLinks":[{"key":"getting-
started"|"linkDisplayName":"Getting started
guide"436|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--
Overview-of-data-world.html"437|"newTab":true438|"svgIcon":"meta"439|"vertically
AlignIcon":true}440|{"key":"documentation"441|"linkDisplayName":"Documentation"4
42|"url":"https:\u002F\u002Fdocs.data.world\u002F"443|"newTab":true444|"svgIcon
":"meta"445|"verticallyAlignIcon":true}446|{"key":"integrations"447|"linkDisplayN
ame":"Integrations"448|"url":"\u002Fintegrations"449|"newTab":true450|"svgIcon":
"integration"451|"verticallyAlignIcon":true452|"showIfAnyDismissed":["onboard-

```

```

checklist"453|"onboard-checklist-admin"]}]454|{"key":"new-
features"455|"linkDisplayName":"New features"456|"url":"https:\u002F\u002Fwhatsn
ew.data.world"457|"newTab":true458|"svgIcon":"magic-
wand"459|"verticallyAlignIcon":true460|"showIfAnyDismissed":["onboard-
checklist"461|"onboard-checklist-
admin"]}]462|{"key":"enterprise"463|"linkDisplayName":"data.world for the
enterprise"464|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-
demo\u002F?src=inapp_data.world-for-the-enterprise"465|"newTab":true466|"svgIcon
":"organization"467|"verticallyAlignIcon":true}]468|"onboardChecklist":{"key":"s
ite"469|"items":[{"id":"joinOrg|"analyticsId":"onboard_checklist.join-org|"hre
f":"\u002Fsettings\u002Forganizations|"completeIfMemberOfOrg":true|"formattedMe
ssage":{"id":"onboard_checklist.join-org|"description":"Text for onboard
checklist join org link|"defaultMessage":"Join organization"}}|{"id":"takeTour"
477|"analyticsId":"onboard_checklist.take_tour"478|"showOnboardModalOnClick":tru
e479|"completeIfDismissed":"onboard-tour-complete"480|"formattedMessage":{"id":"
onboard_checklist.take_tour"481|"description":"Text for onboard checklist take
tour link"482|"defaultMessage":"Take the tour"}}483|{"id":"customizeProfile"484|
"analyticsId":"onboard_checklist.customize_profile"485|"href":"\u002Fsettings"48
6|"dismissOnClick":"customize-profile-link"487|"completeIfDismissed":"customize-
profile-link"488|"formattedMessage":{"id":"onboard_checklist.customize_profile"4
89|"description":"Text for onboard checklist customize profile
link"490|"defaultMessage":"Upload your
photo"}}491|{"id":"createProject"492|"analyticsId":"onboard_checklist.create-
project"493|"href":"\u002Fcreate-a-project"494|"dismissOnClick":"create-project-
link"495|"completeIfDismissed":"create-project-
link"496|"formattedMessage":{"id":"onboard_checklist.create-
project"497|"description":"Text for onboard checklist create project
link"498|"defaultMessage":"Create your first project"}}499|{"id":"exploreIntegra
tions"500|"analyticsId":"onboard_checklist.explore-integrations"501|"href":"\u00
2Fintegrations"502|"target":"_blank"503|"dismissOnClick":"explore-integrations-
link"504|"completeIfDismissed":"explore-integrations-
link"505|"formattedMessage":{"id":"onboard_checklist.explore-
integrations"506|"description":"Text for onboard checklist explore integrations
link"507|"defaultMessage":"Explore
integrations"}}}]508|"onboardTour":[{"show":true509|"header":"Welcome to
data.world!"510|"subhead":"The place for modern data
teamwork."511|"leftCol":"Organize data work so you and your team find what you
need&nbsp;faster."512|"middleCol":"Explore and create a shared understanding of
your&nbsp;data."513|"rightCol":"Share insights and invite more people to
participate in data&nbsp;teamwork."}]514|{"show":true515|"imageHeaderKey":"search
"516|"imageKey":"collaborate"517|"imageAlt":"Start with Data"518|"header":"Start
with fantastic&nbsp;data"519|"subhead":"Browse thousands of free520| open data
resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from
almost anywhere into a single searchable&nbsp;library.\n\nMake it open521| share
with your team522| or keep it&nbsp;private."}523|{"show":true524|"imageHeaderKey
":"bulb"525|"imageKey":"dataProject"526|"imageAlt":"Data
Projects"527|"header":"Organize your data&nbsp;projects"528|"subhead":"Create
projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd

```

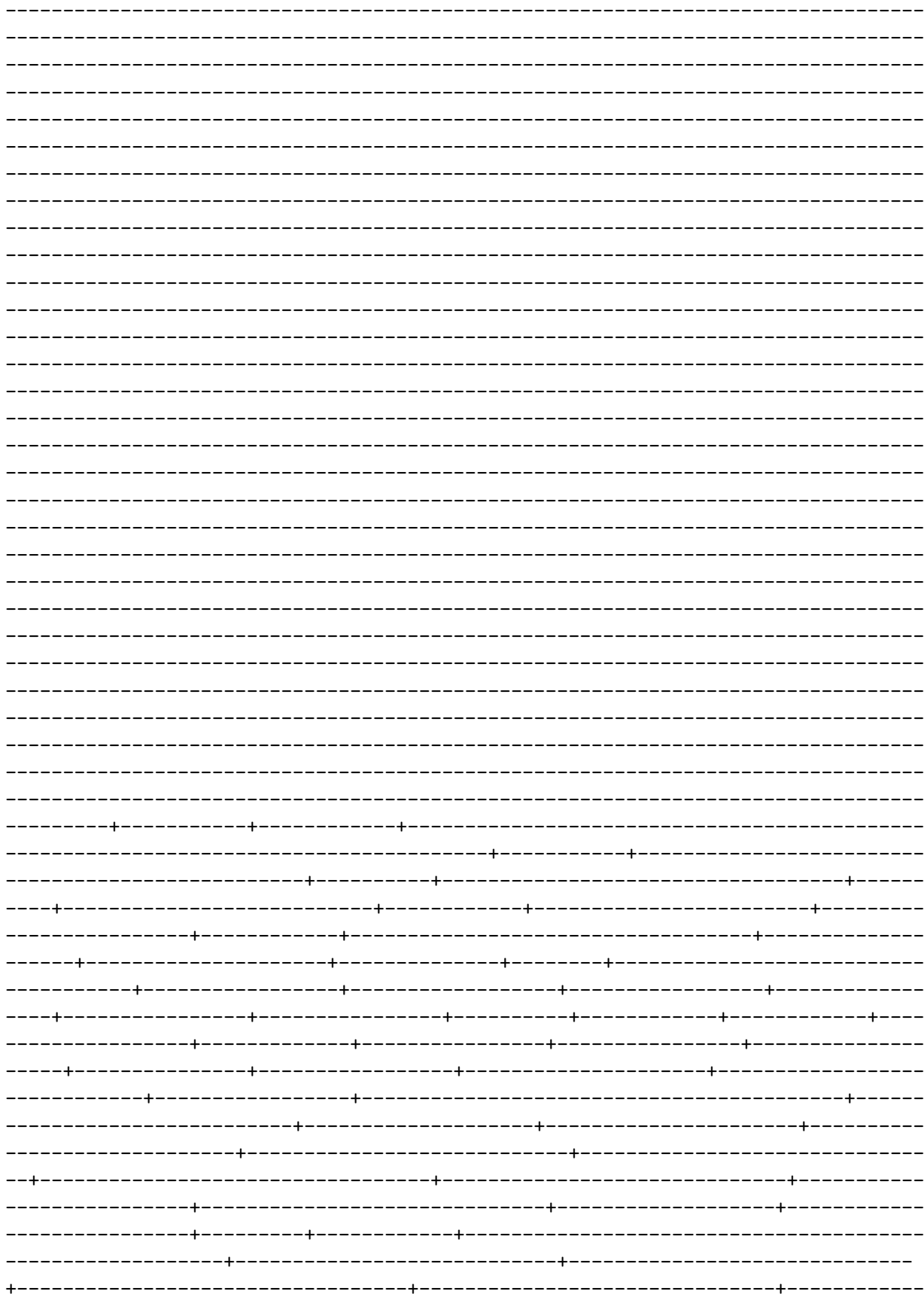
context to make it easier for others to discover529| understand530| and contribute to your projects.\n\nImmediately query data without leaving your projects."}531|{"show":true532|"imageHeaderKey":"code"533|"imageKey":"slideIntegrations"534|"imageAlt":"Integrations"535|"reverse":true536|"header":"Connect the data tools you love"537|"subhead":"Automatically keep data.world synced with your data sources.\n\nIntegrate your team's favorite tools538| from analysis to visualization."}539|{"show":true540|"imageHeaderKey":"discussions"541|"imageKey":"discussionsCollaboration"542|"imageAlt":"Contributors"543|"header":"Add more brainpower to your projects"544|"subhead":"Invite your team to securely and effectively work together.\n\nManage fine-grained permissions for each member.\n\nAsk questions545| add insights546| bookmark top contributions547| and more."}548|{"show":true549|"imageHeaderKey":"insight"550|"imageKey":"knowYourData"551|"imageAlt":"know your data"552|"header":"Share your results"553|"subhead":"Share meaningful insights and the data behind them.\n\nDistribute clean554| documented data for others to use in their own work.\n\nHelp the right people discover the resources they need."}]]555|"other":{"resourceLinks":[{"key":"getting-started|"linkDisplayName":"Getting started guide"557|"url":"https:\u002F\u002Fdocs.data.world\u002Fen\u002F59261-60585-0--Overview-of-data-world.html"558|"newTab":true559|"svgIcon":"meta"560|"verticallyAlignIcon":true}561|{"key":"documentation"562|"linkDisplayName":"Documentation"563|"url":"https:\u002F\u002Fdocs.data.world\u002F"564|"newTab":true565|"svgIcon":"meta"566|"verticallyAlignIcon":true}567|{"key":"integrations"568|"linkDisplayName":"Integrations"569|"url":"\u002Fintegrations"570|"newTab":true571|"svgIcon":"integration"572|"verticallyAlignIcon":true573|"showIfAnyDismissed":["onboard-checklist"574|"onboard-checklist-admin"]}575|{"key":"new-features"576|"linkDisplayName":"New features"577|"url":"https:\u002F\u002Fwhatsnew.data.world"578|"newTab":true579|"svgIcon":"magic-wand"580|"verticallyAlignIcon":true581|"showIfAnyDismissed":["onboard-checklist"582|"onboard-checklist-admin"]}583|{"key":"enterprise"584|"linkDisplayName":"data.world for the enterprise"585|"url":"https:\u002F\u002Fdata.world\u002Fschedule-a-demo\u002F?src=inapp_data.world-for-the-enterprise"586|"newTab":true587|"svgIcon":"organization"588|"verticallyAlignIcon":true}589|"enableImageUpload":true|"onboardChecklist":{"key":"site"591|"items":[{"id":"createAccount"592|"analyticsId":"onboard_checklist.create_account"593|"complete":true594|"formattedMessage":{"id":"onboard_checklist.create_account"595|"description":"Text for onboard checklist create account link"596|"defaultMessage":"Create account"}}597|{"id":"takeTour"598|"analyticsId":"onboard_checklist.take_tour"599|"showOnboardModalOnClick":true600|"completeIfDismissed":"onboard-tour-complete"601|"formattedMessage":{"id":"onboard_checklist.take_tour"602|"description":"Text for onboard checklist take tour link"603|"defaultMessage":"Take the tour"}}604|{"id":"customizeProfile"605|"analyticsId":"onboard_checklist.customize_profile"606|"href":"\u002Fsettings"607|"dismissOnClick":"customize-profile-link"608|"completeIfDismissed":"customize-profile-link"609|"formattedMessage":{"id":"onboard_checklist.customize_profile"610|"description":"Text for onboard checklist customize profile link"611|"defaultMessage":"Upload your photo"}}612|{"id":"createProject"613|"analyticsId":"onboard_checklist.create-

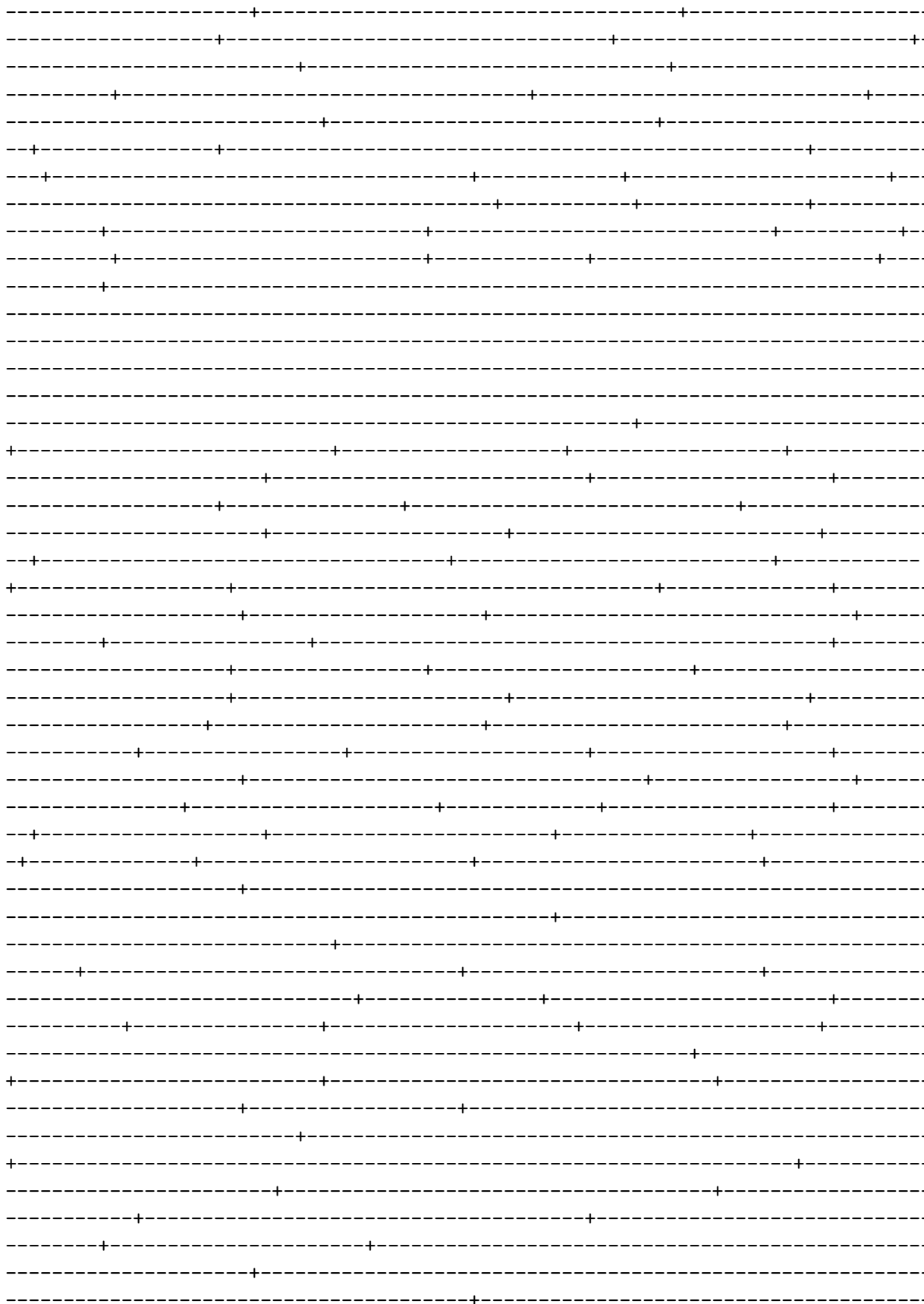
```

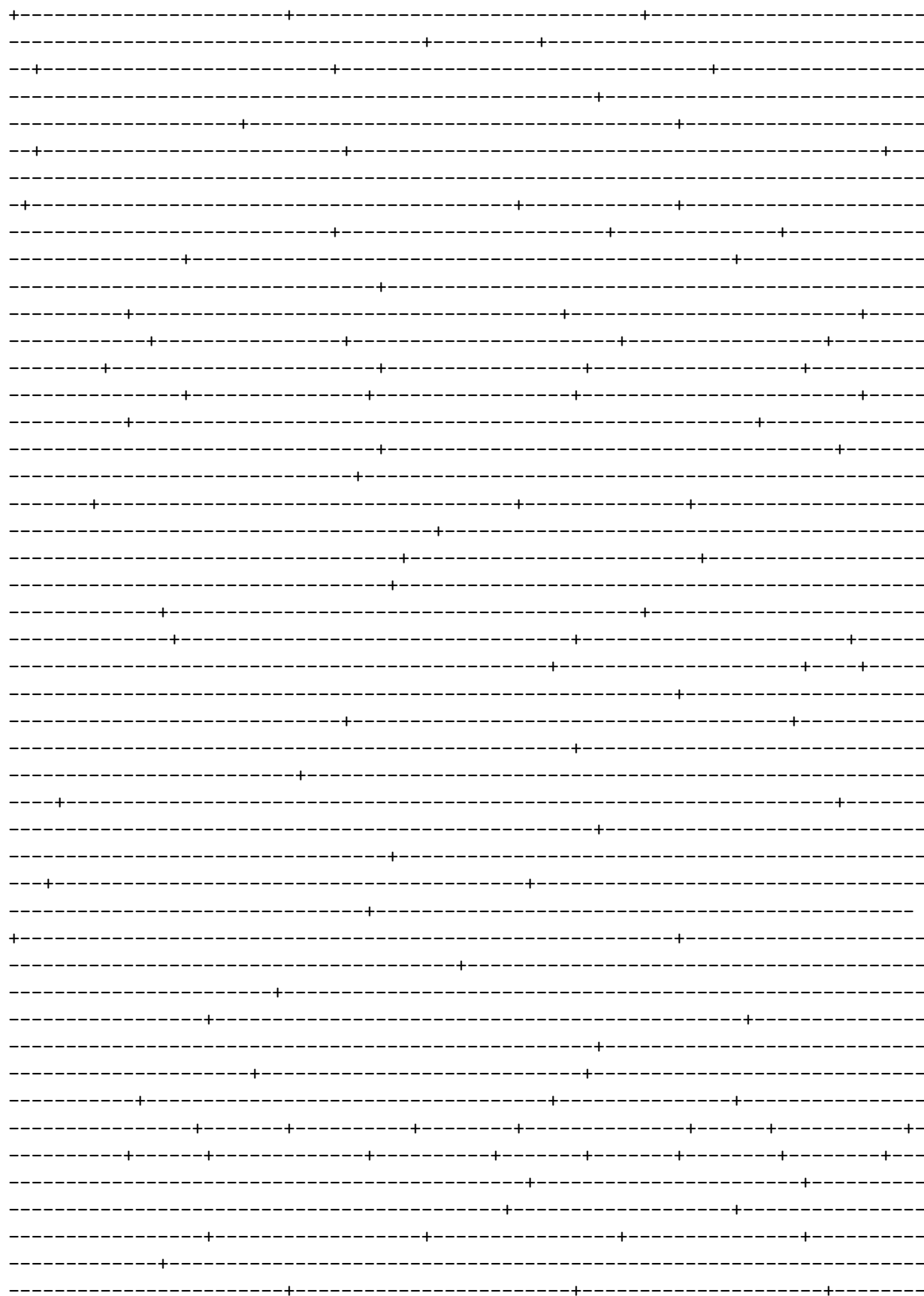
project"614|"href": "\u002Fcreate-a-project"615|"dismissOnClick": "create-project-
link"616|"completeIfDismissed": "create-project-
link"617|"formattedMessage": {"id": "onboard_checklist.create-
project"618|"description": "Text for onboard checklist create project
link"619|"defaultMessage": "Create your first project"}}620|{"id": "exploreIntegra
tions"621|"analyticsId": "onboard_checklist.explore-integrations"622|"href": "\u00
2Fintegrations"623|"target": "_blank"624|"dismissOnClick": "explore-integrations-
link"625|"completeIfDismissed": "explore-integrations-
link"626|"formattedMessage": {"id": "onboard_checklist.explore-
integrations"627|"description": "Text for onboard checklist explore integrations
link"628|"defaultMessage": "Explore
integrations"}}}]629|"onboardTour": [{"show": true630|"header": "Welcome to
data.world!"631|"subhead": "The place for modern data
teamwork."632|"leftCol": "Organize data work so you and your team find what you
need&nbsp;faster."633|"middleCol": "Explore and create a shared understanding of
your&nbsp;data."634|"rightCol": "Share insights and invite more people to
participate in data&nbsp;teamwork."}635|{"show": true636|"imageHeaderKey": "search
"637|"imageKey": "collaborate"638|"imageAlt": "Start with Data"639|"header": "Start
with fantastic&nbsp;data"640|"subhead": "Browse thousands of free641| open data
resources; bookmark your favorites for&nbsp;later.\n\nGather your own data from
almost anywhere into a single searchable&nbsp;library.\n\nMake it open642| share
with your team643| or keep it&nbsp;private."}644|{"show": true645|"imageHeaderKey
": "bulb"646|"imageKey": "dataProject"647|"imageAlt": "Data
Projects"648|"header": "Organize your data&nbsp;projects"649|"subhead": "Create
projects that hold everything you need to answer critical&nbsp;questions.\n\nAdd
context to make it easier for others to discover650| understand651| and
contribute to your&nbsp;projects.\n\nImmediately query data without leaving your
&nbsp;projects."}652|{"show": true653|"imageHeaderKey": "code"654|"imageKey": "slid
eIntegrations"655|"imageAlt": "Integrations"656|"reverse": true657|"header": "Conne
ct the data tools you&nbsp;love"658|"subhead": "Automatically keep data.world
synced with your data&nbsp;sources.\n\nIntegrate your team's favorite tools659|
from analysis to&nbsp;visualization."}660|{"show": true661|"imageHeaderKey": "disc
ussions"662|"imageKey": "discussionsCollaboration"663|"imageAlt": "Contributors"66
4|"header": "Add more brainpower to&nbsp;your&nbsp;projects"665|"subhead": "Invite
your team to securely and effectively work&nbsp;together.\n\nManage fine-grained
permissions for each&nbsp;member.\n\nAsk questions666| add insights667| bookmark
top contributions668| and&nbsp;more."}669|{"show": true670|"imageHeaderKey": "insi
ght"671|"imageKey": "knowYourData"672|"imageAlt": "know your
data"673|"header": "Share your results"674|"subhead": "Share meaningful insights
and the data behind&nbsp;them.\n\nDistribute clean675| documented data for
others to use in their own&nbsp;work.\n\nHelp the right people discover the
resources they&nbsp;need."}]676|"newResourceItems": [{"type": "dataset"| "exampleL
ink": "https:\u002F\u002Fdata.world\u002Fninja\u002Fanw-obstacle-history"}|{"type
": "project"| "exampleLink": "https:\u002F\u002Fdata.world\u002Fmarkmarkoh\u002Fexp
loring-thor"}]]|"version_build": "20230921123426_7d6ba6c4"| "version_check_interva
l": 60000|"version_commit": "7d6ba6c4"| "visibility": {"discoverableVisibilitySubHea
derText": "This \u003C%= resourceType %\u003E can be discovered by all data.world
users. You can change the \u003C%= resourceType %\u003E visibility or make it

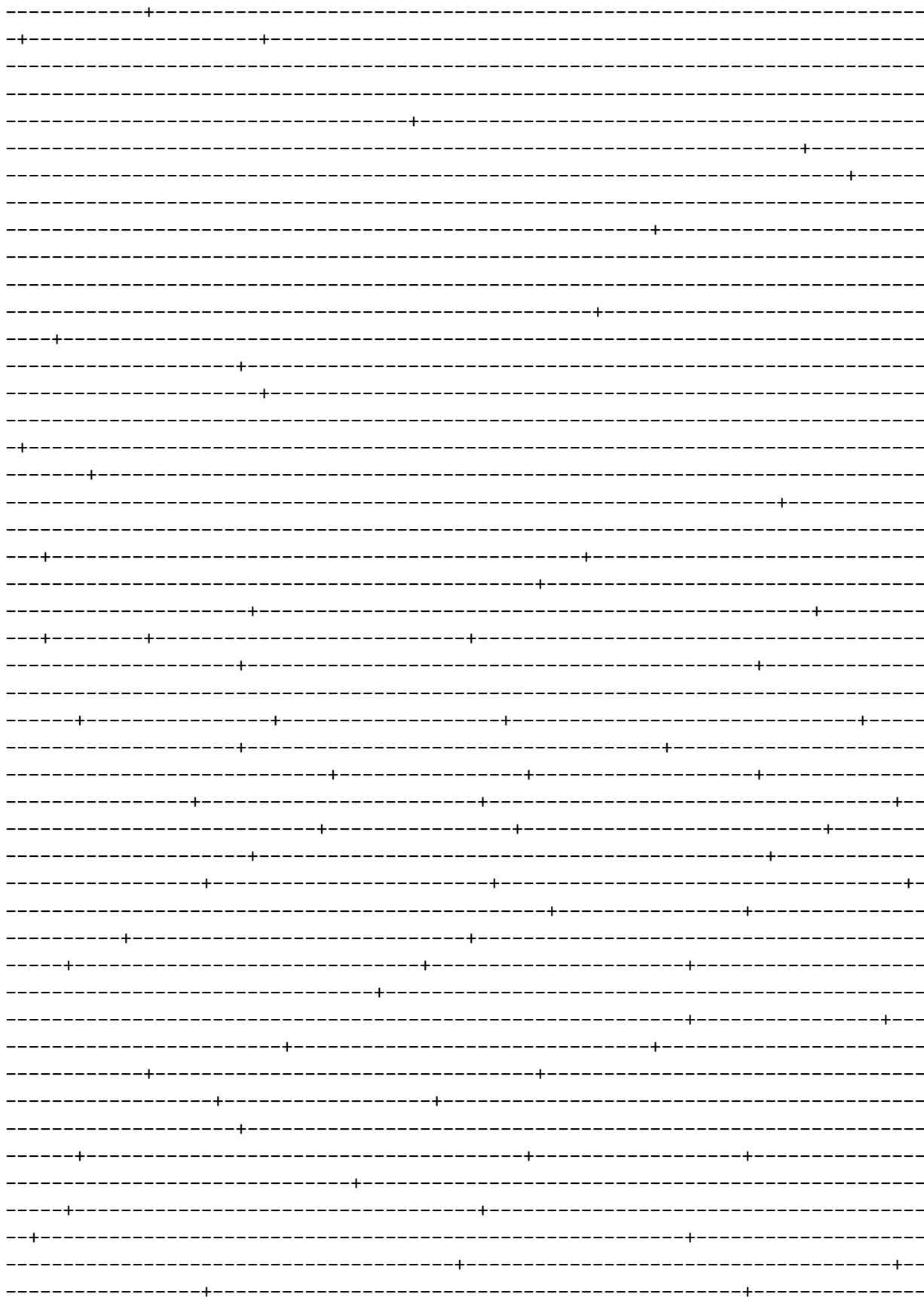
```

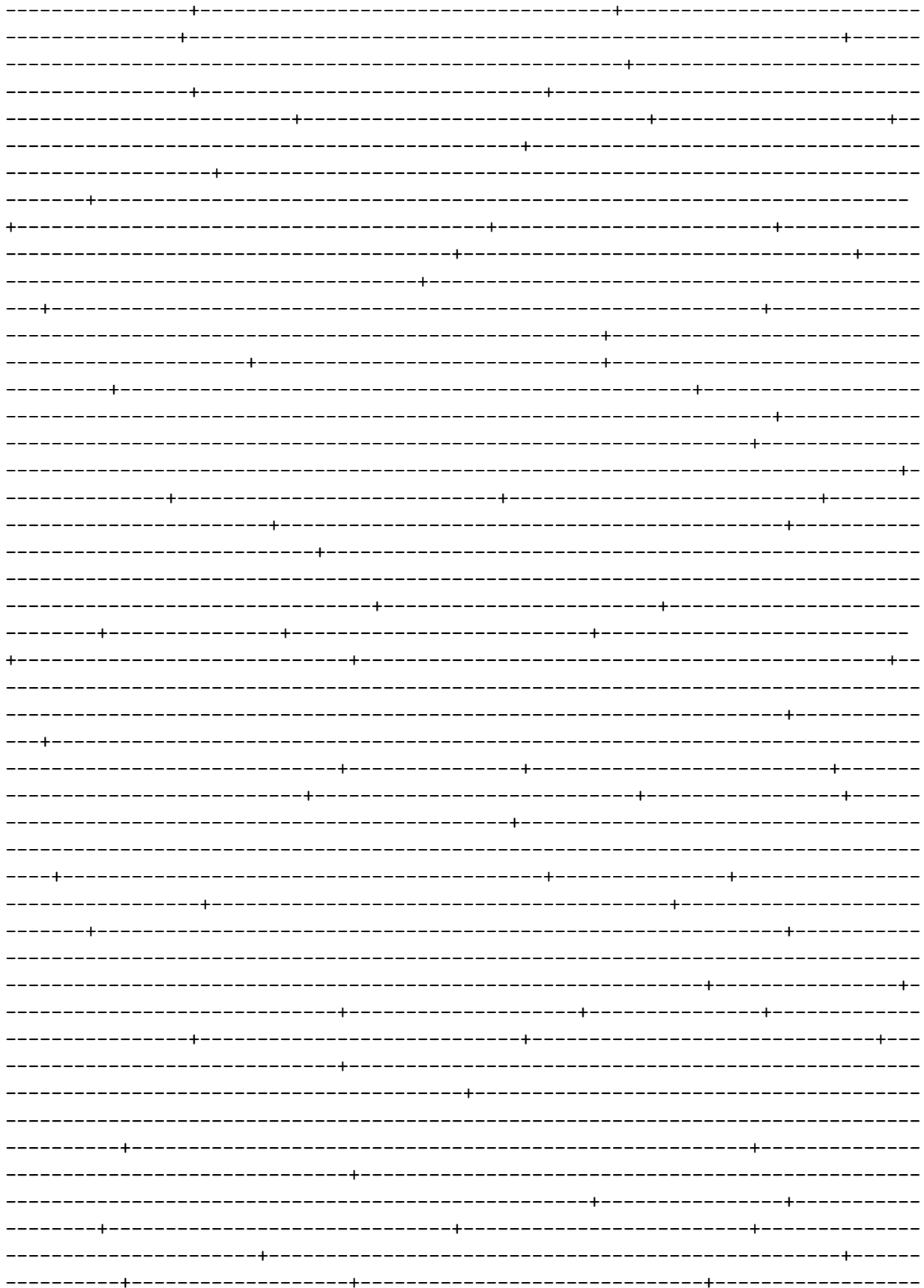
```
private in the table below."|"openVisibilitySubHeaderText": "This \u003C%= resourceType %\u003E can be discovered and viewed by anyone on data.world. You can change the \u003C%= resourceType %\u003E visibility or make it private in the table below."|"privateVisibilityShareButtonText": "Share with community"|"privateVisibilitySubHeaderText": "This \u003C%= resourceType %\u003E is only visible to you. It is neither discoverable nor viewable by the data.world community. To open it up to others to view| change the visibility settings below."|"shareModalBodyText": "Set the visibility level for all of data.world below. The content of the \u003C%= resourceType %\u003E will only be visible to those who have been granted explicit access to this \u003C%= resourceType %\u003E."|"shareModalHeaderText": "Share with all data.world"|"visibilityTableTitleCellText": "Anyone on data.world"}|"context_id": "b04792da5b608cdd"}</script><script type="application/json" id="init-profile">{"requestAccessForm": [{"label": "Full name"|"required": true693|"readOnly": true694|"showOptionalFlag": false695|"lookupProperty": "displayName"696|"type": "input"}] [{"label": "Contact email"|"placeholder": "Provide an email in the case a dataset administrator needs to reach out"699|"required": false|"readOnly": false701|"showOptionalFlag": true|"type": "email"}] [{"label": "Message"|"placeholder": "Share what you would like to do with this dataset"705|"required": true706|"readOnly": false707|"showOptionalFlag": false708|"type": "textarea"}]}</script><script type="application/json" id="init-analytics">{"siteId": "00000"}</script><script src="https://js.stripe.com/v3/" async id="stripe-js"></script> <script type="text/javascript" src="https://assets.data.world/js/runtime.4c0777f218351a402063.js"></script><script type="text/javascript" src="https://assets.data.world/js/vendors~app.73c77cf63888ff69bcb2.js"></script><script type="text/javascript" src="https://assets.data.world/js/vendors~app-vendors.66b69a1400c3cdba7c1a.js"></script><script type="text/javascript" src="https://assets.data.world/js/vendors~vendors.595dac19a1bcced9e705.js"></script><script type="text/javascript" src="https://assets.data.world/js/vendors.6dfd2438038a5ca9e317.js"></script><script type="text/javascript" src="https://assets.data.world/js/app.15687b4170eaf580bb02.js"></script></body></html>|
```

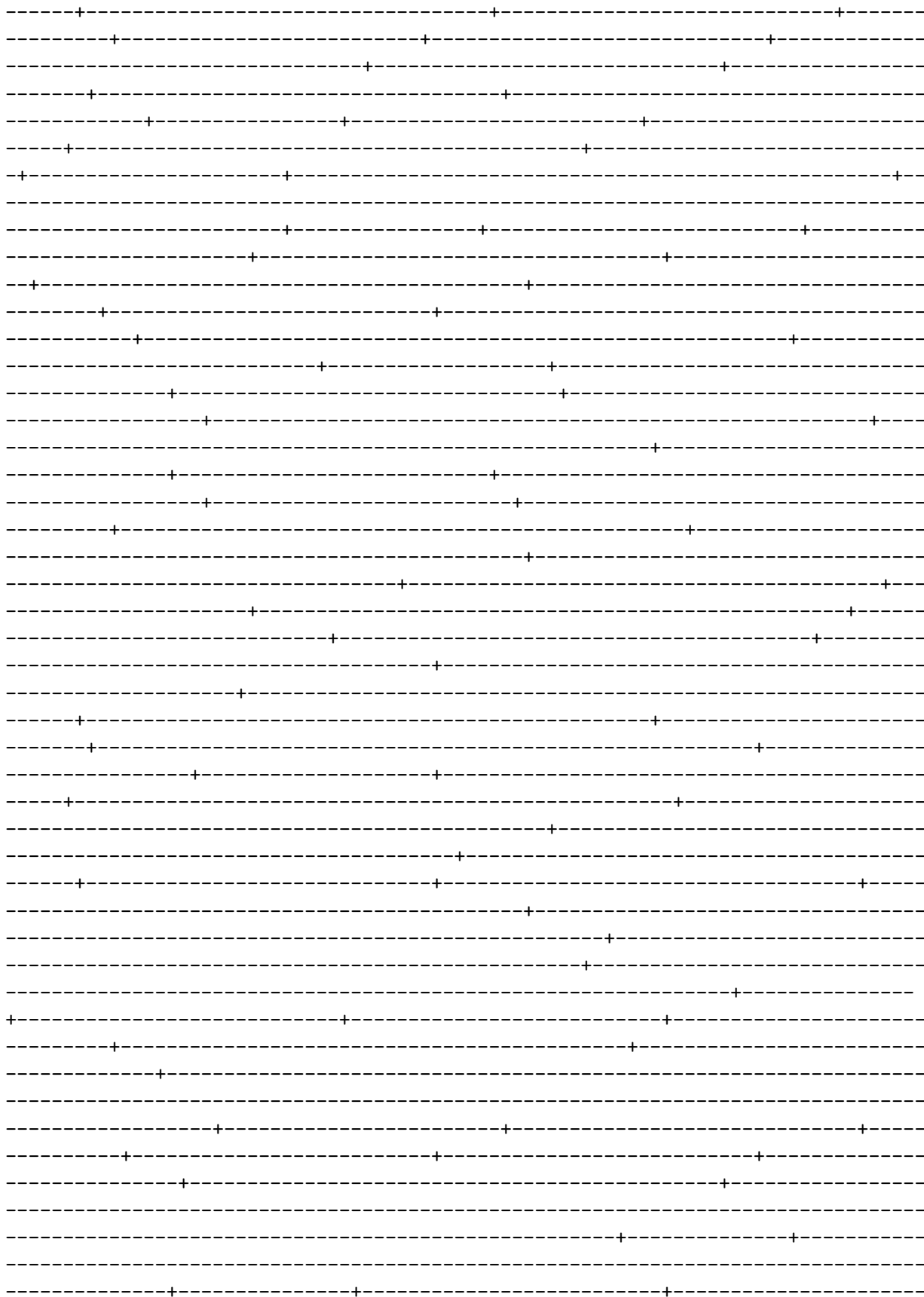


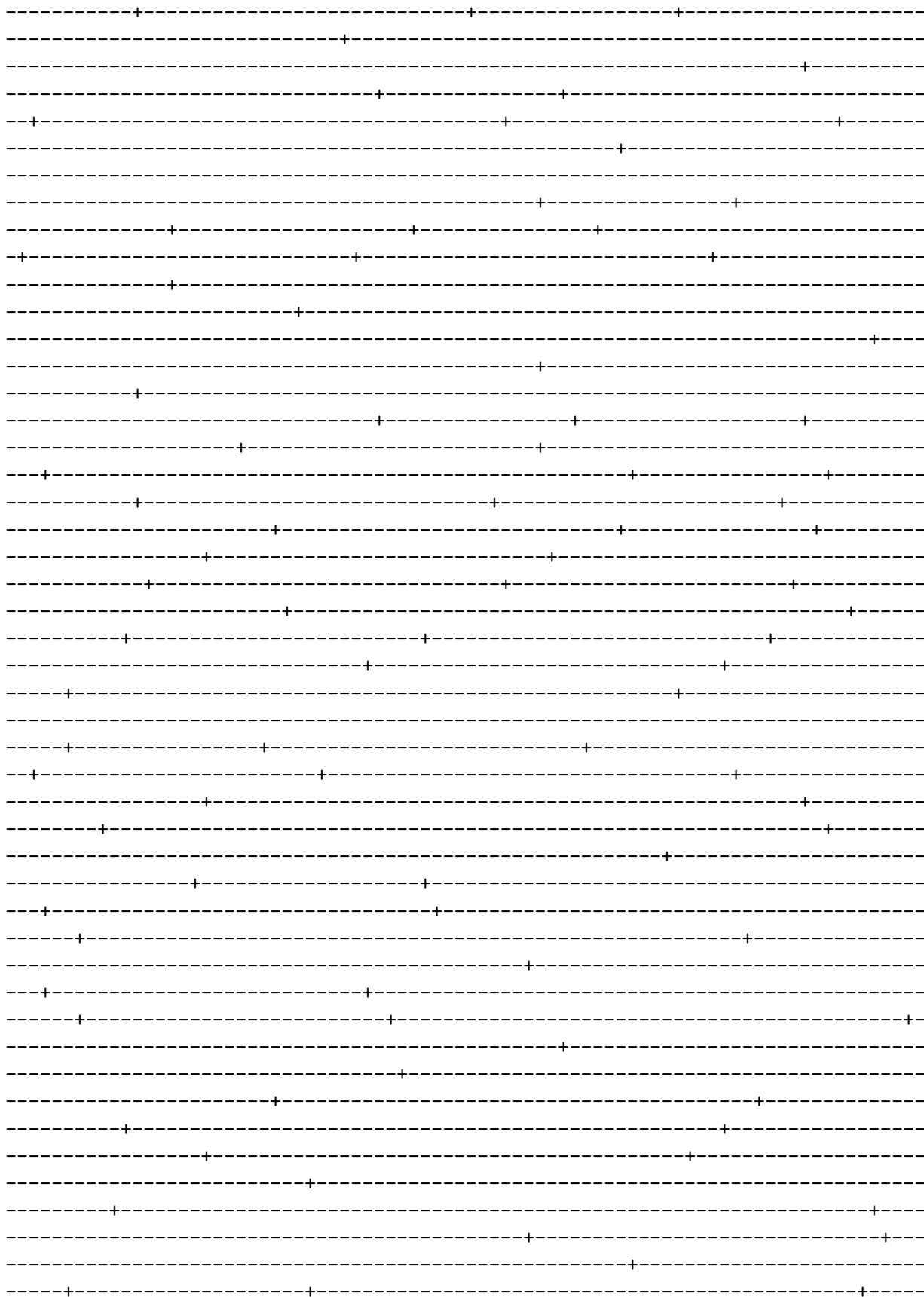


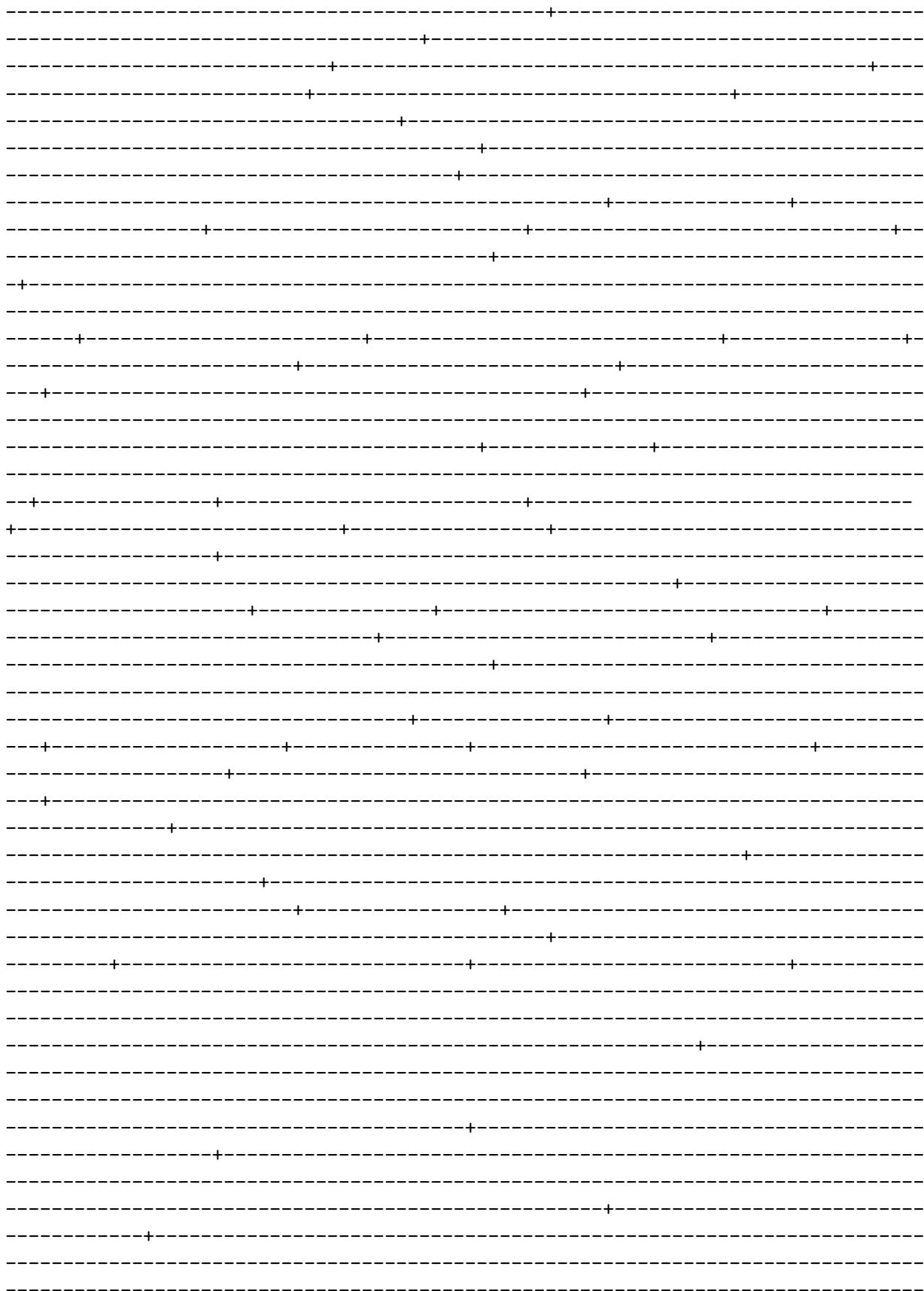












[illegible]

[illegible]

A 40x40 grid of 1600 '0' and '1' characters. The '1's are sparsely distributed across the grid, forming a complex, abstract pattern. The pattern is not symmetric and appears to be a random or pseudo-random distribution of the two digits.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

null		null		null	
null		null		null	
null			null		null
null		null			null
null		null			
null			null		
null	null		null		
null		null		null	
null			null		null
null				null	null
null	null			null	
null			null		null
null				null	
null		null			
null		null			null
null			null		null
null				null	
null					null
null					
null		null			
null		null		null	
null			null		null
null				null	
null					null
null		null			null
null			null		
null				null	
null					null
null	null		null		null
null			null		
null		null		null	
null		null			null
null			null		null
null				null	
null					null
null	null		null		null
null		null		null	
null			null		null
null				null	
null					null
null		null		null	
null		null			
null			null		
null	null				
null			null		null
null				null	

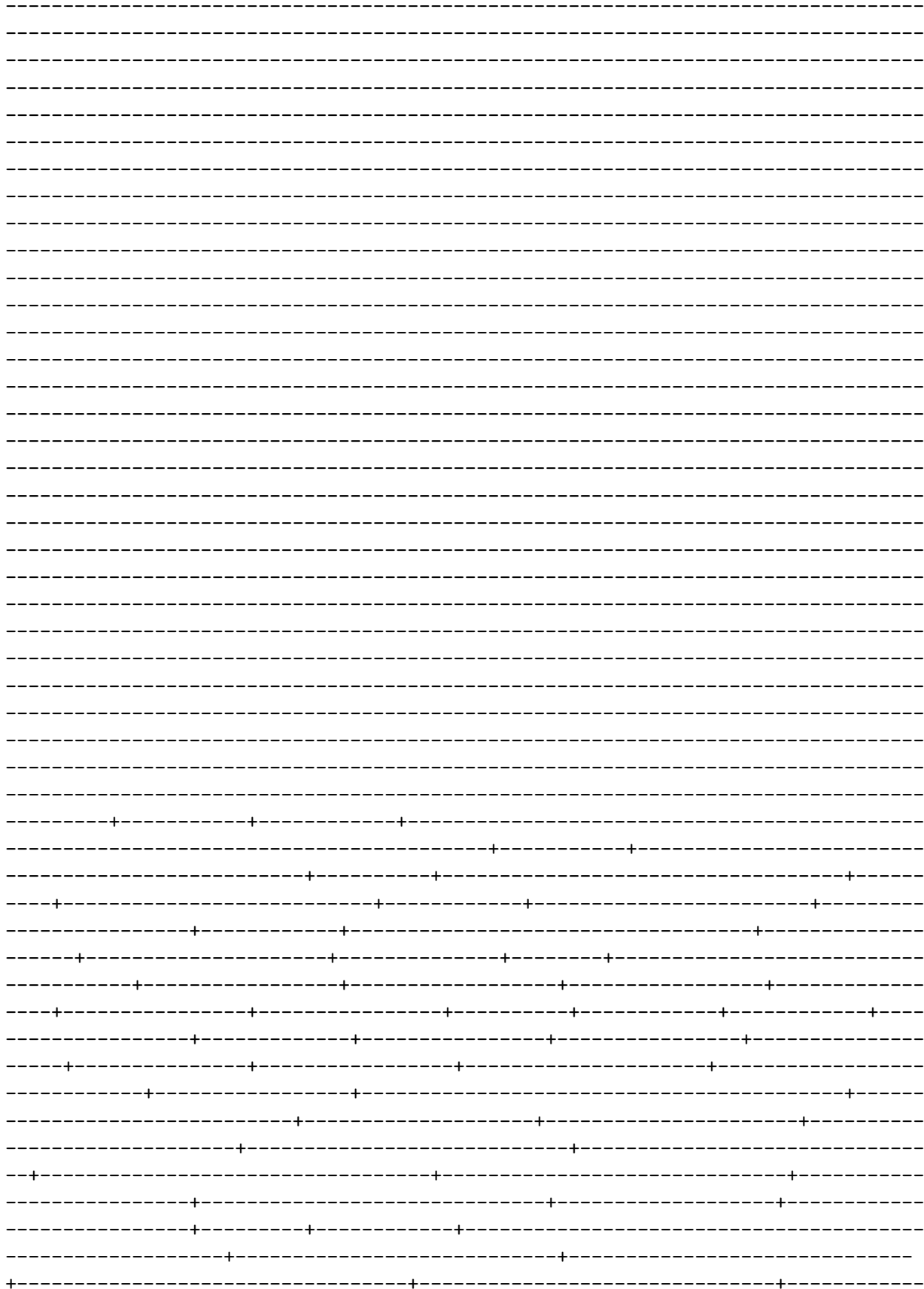
[illegible]

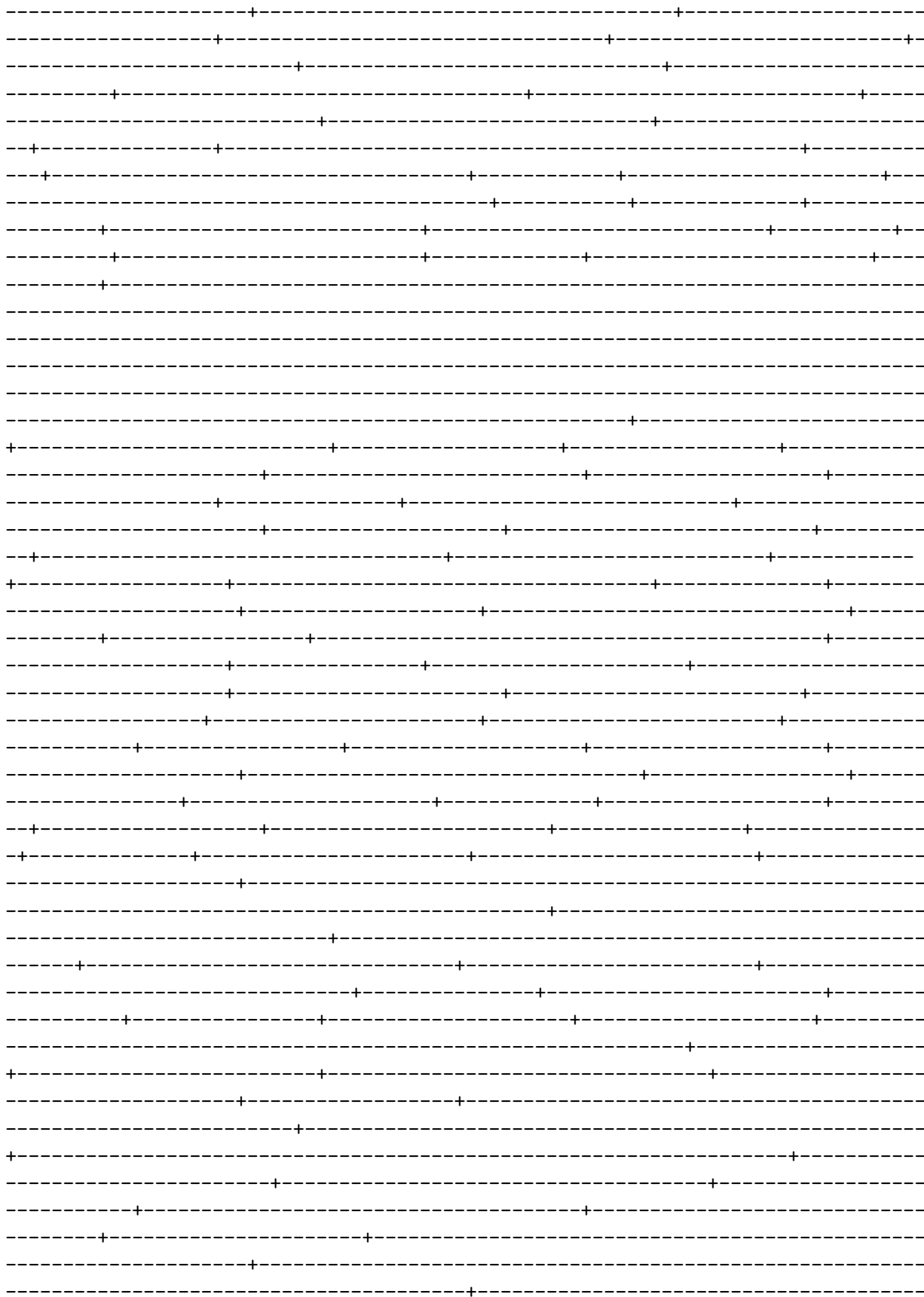
null		null			
null			null		
null				null	
null					null
null		null			null
null					
null					
null					
null	null		null		null
null			null		
null		null		null	null
null		null		null	
null	null				
null	null		null		
null		null		null	
null			null		null
null			null		
null	null		null		null
null		null		null	
null			null		null
null			null		
null				null	
null	null				null
null		null		null	
null			null		null
null				null	
null					null
null					
null	null		null		null
null		null		null	
null	null		null		null
null		null		null	
null		null			null
null		null		null	
null			null		null
null				null	
null					null
null					
null	null		null		null
null		null		null	
null			null		null
null				null	
null	null				null
null			null		
null				null	
null		null			null
null		null			
null			null		null
null				null	
null					null
null					

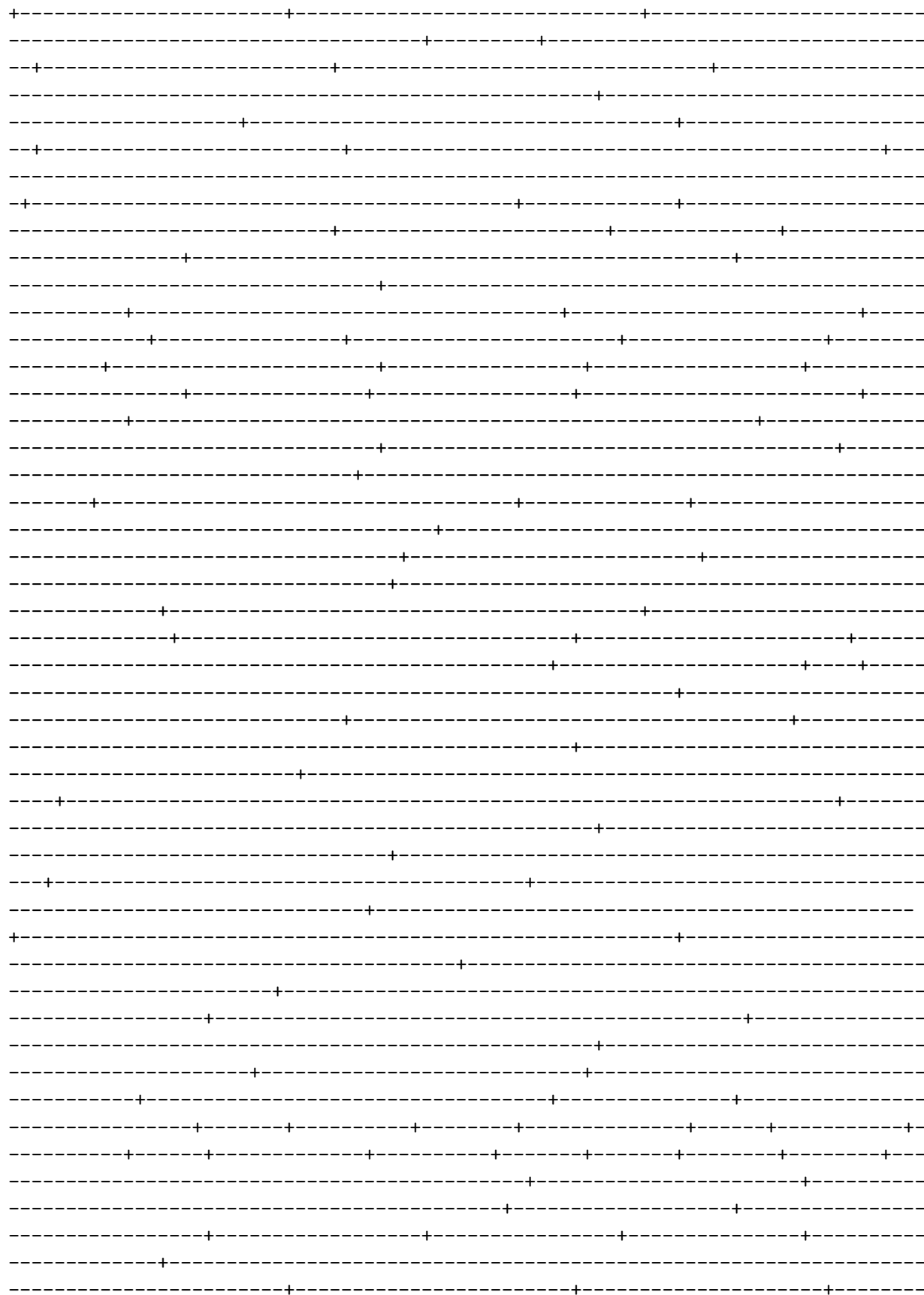
null		null		null
null				null
null			null	null
null		null		
null			null	
null				null
null		null		
null		null	null	
null			null	
null		null		null
null				null
null	null		null	null
null			null	
null		null		null
null		null	null	
null	null			null
null	null		null	
null		null		null
null			null	
null		null	null	
null				null
null		null		null
null	null		null	null
null		null		null
null			null	
null		null		null
null	null			null
null		null	null	
null				null
null		null		null
null			null	
null		null		null
null	null			null
null	null	null		null
null	null		null	
null	null	null		null
null	null			null
null	null	null		null
null	null		null	
null	null	null		null
null	null			null
null	max			
null				

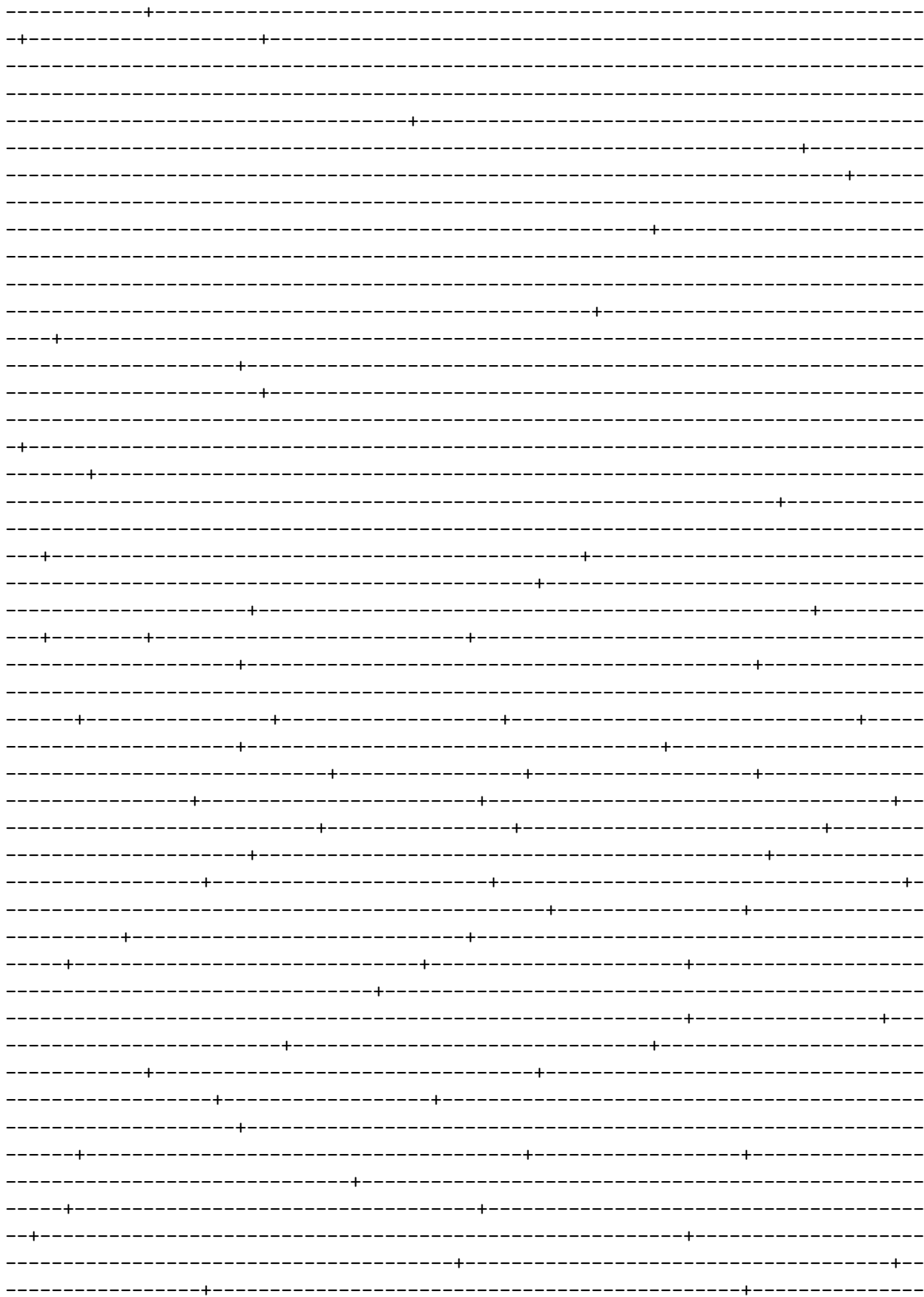
[illegible]

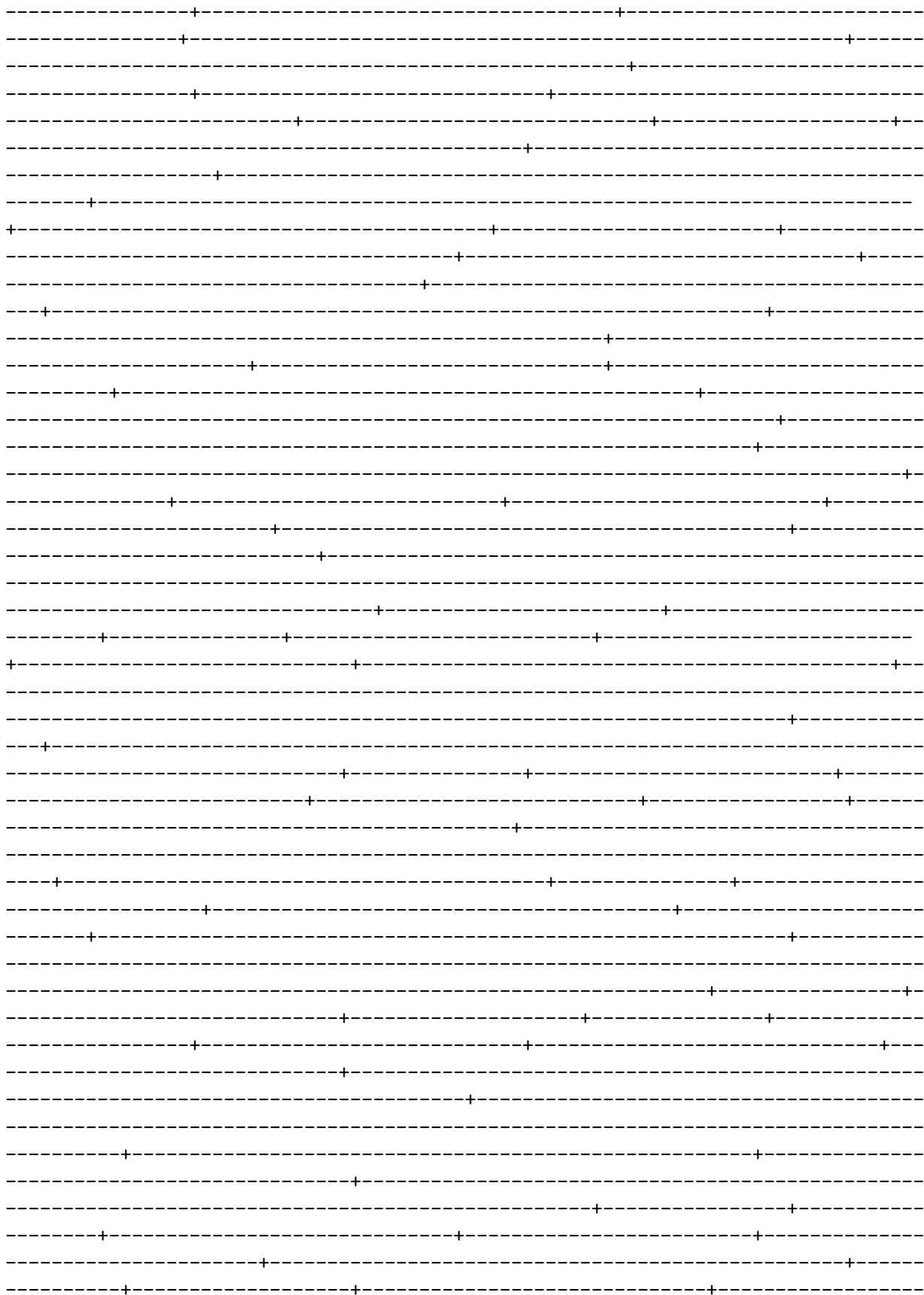
[illegible]

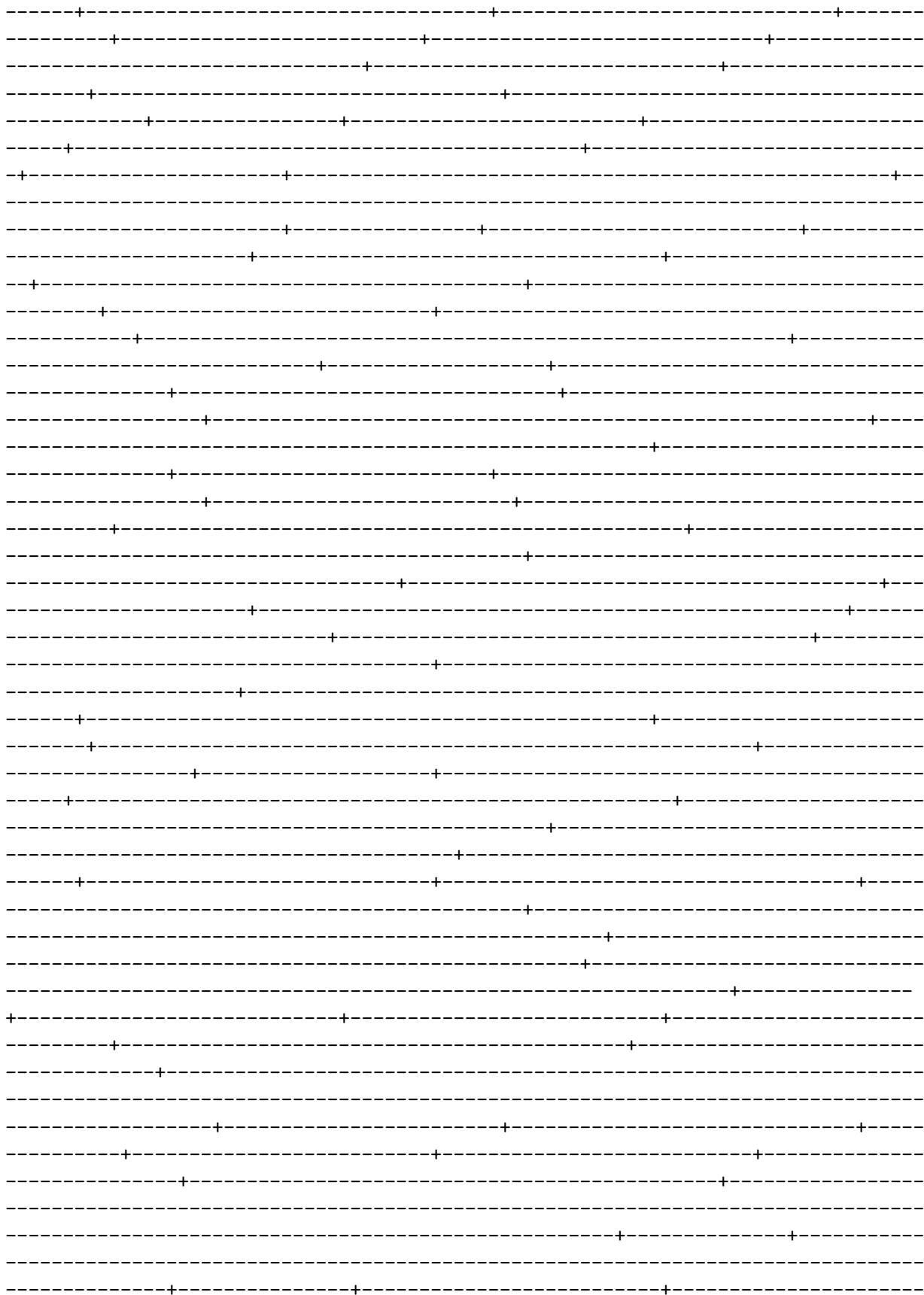


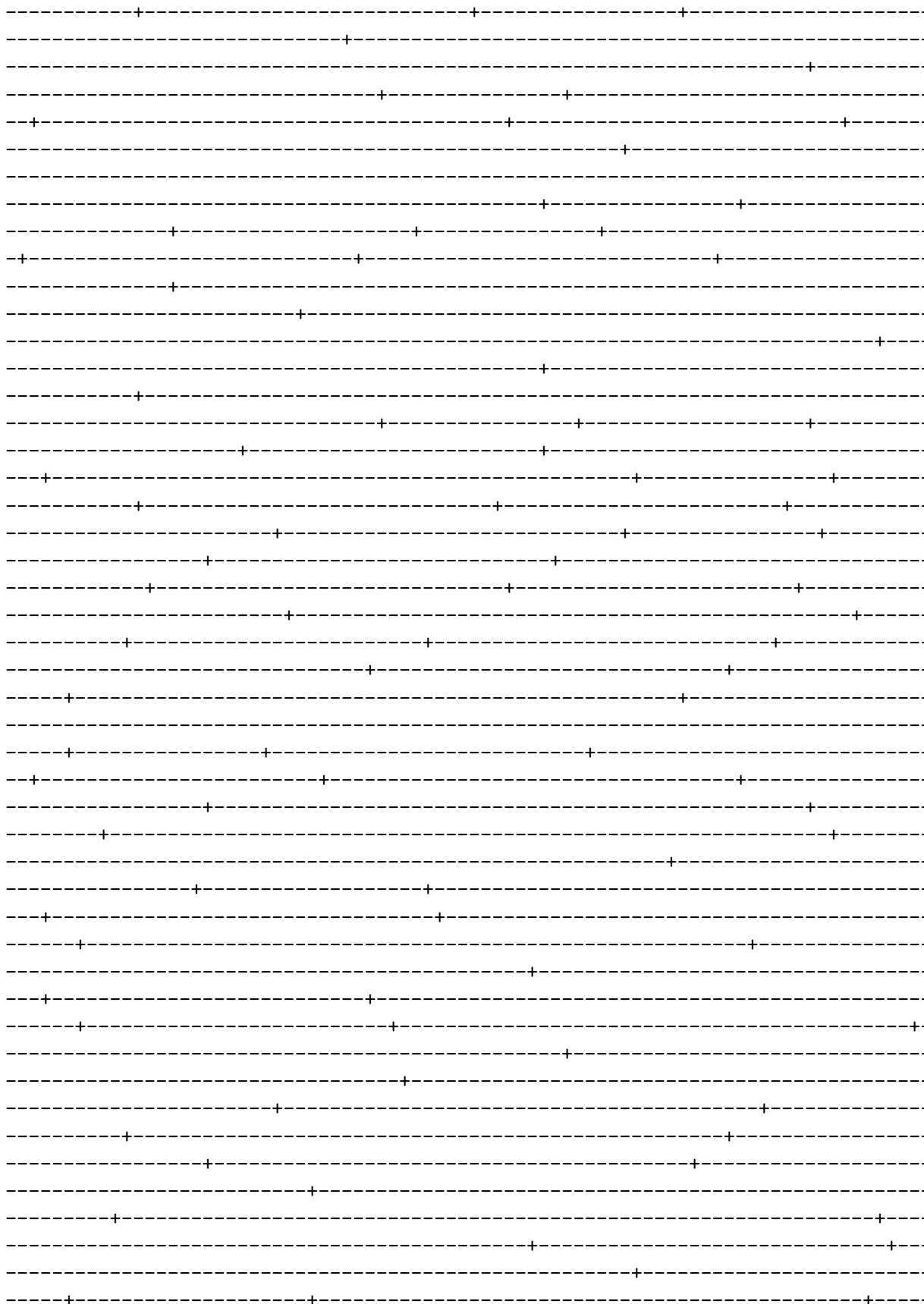


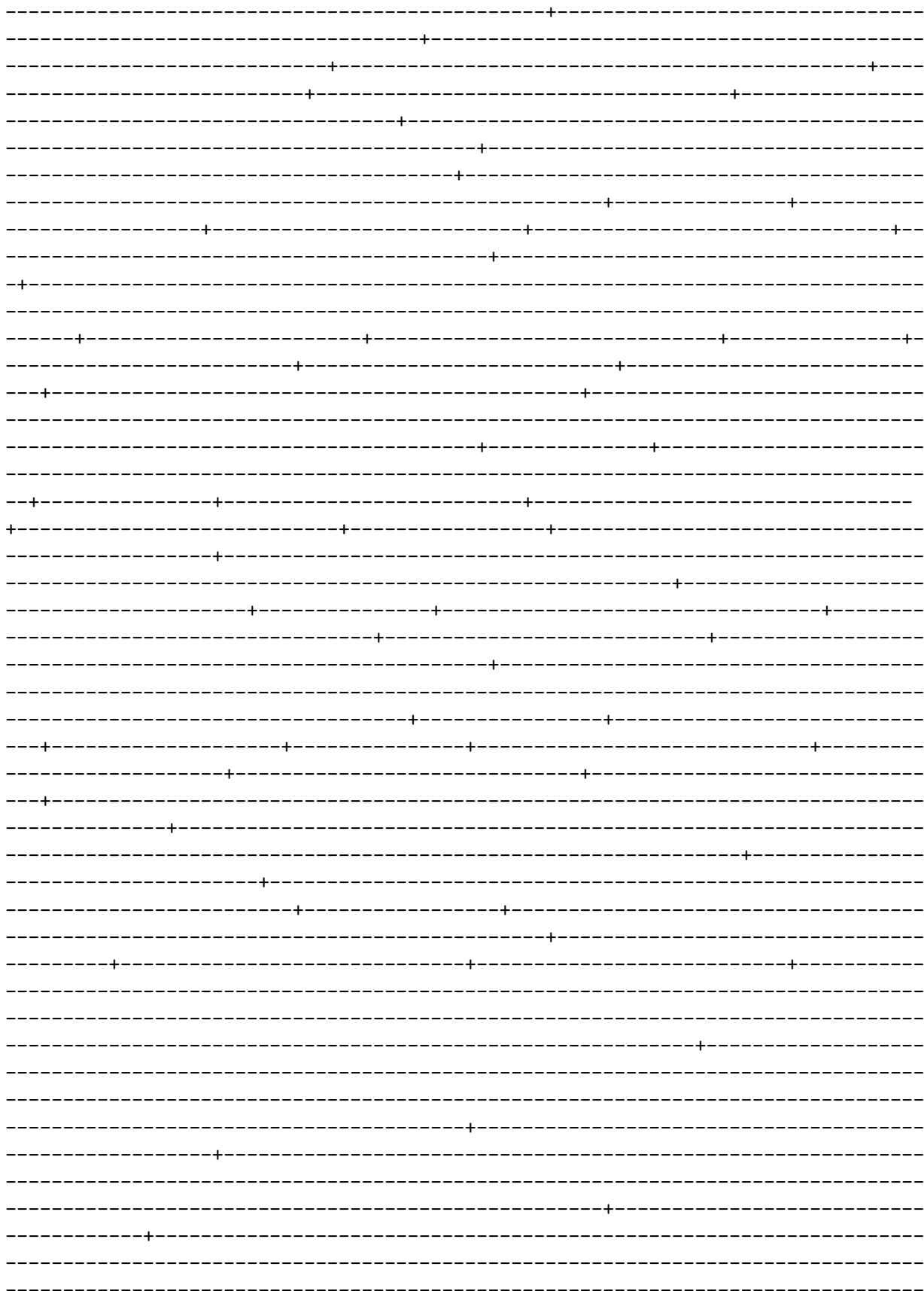












```
[39]: sub_nd_british2_df.write.save("/tmp/dcd/pyspark/british2/", format='csv',  
    ↪ header=True)
```

```
[ ]: sub_nd_british2_df.write.save("gs://dcd05omg-bucket/dcd/pyspark/british2/",  
    ↪format='csv', header=True)
```