UKCRAWL: CODE NOTES AND SNIPPETS

The parquet format has several benefits over alternatives like CSV; parquet supports file compression, allowing for smaller file sizes, and data types are preserved, meaning the 'all_postcodes' list column is parsed correctly.

Most major programming languages have libraries that can read parquet files. The following examples show how this data may be read.

In Python there are a number of libraries that can read parquet files, including `pandas`, `polars`, and `duckdb`:

```python
import pandas as pd
import polars as pl
import duckdb

df = pd.read_parquet("ukcrawl-2019.parquet")
df = pl.read_parquet("ukcrawl-2019.parquet")
df = duckdb.read_parquet("ukcrawl-2019.parquet")
```

In R you can use the `arrow` library (https://arrow.apache.org/docs/r/):

```r
library(arrow)

# reads into a 'tibble'
df <- read_parquet("ukcrawl-2019.parquet")
```