**Project Overview**

Background

Luckin Coffee is a Chinese company which was listed on Nasdaq on May 16th, 2019. Early this year, Muddy Waters revealed a report which claimed that Luckin Coffee was inflating its number of items sold per store. After Lucking announced on April 2 that it would conduct an internal investigation into fraud allegations. Luckin’s stock price dropped more than 80%. This event aroused our interests. For this project, we decided to use Luckin’s earning announcements to develop a graphical database to apply the model to automated predictive text completion.

Approach

Our intention was to create a graph database based on the earnings statements for mid-term submission but due to issues with Neo4j we focused creating a predictive text filler through python coding. For the final assignment submission, we intend to make modifications to this code to allow it to be integrated into Neo4j to create a graph database.

Methodology

Data: Luckin Coffee Earnings Call Transcript was attained online and converted into a txt file for ingestion

NLTK: Python’s built-in programs to work with language data. It provides us with sample texts and tutorial code.

N-gram model: the model that can assign probabilities to the sequences of words. It allows our program to output the predicted text.

Neo4j: the graph database software. It can help us to a build graph database which can be connected to python programs by API.

Mid-term output

A python program creates a dictionary of n-gram (set to n=5 for this demonstration) strings and prompts the user to enter text based on these n-grams. Once submitted, the code will automatically complete the remainder of the paragraph based on the text from the earnings statement. For the final assignment, we will integrate this into a graphical database created by Neo4j. The assignment output has been uploaded to our respective github sites the links to which can be found in the canvas submission.