**A Study on cloud platforms and Analytics assessment**

Table of Contents

[Question 1: Build a Chatbot 3](#_Toc181020528)

[1.1: Five Functionalities Related to Tourism for the Chatbot 3](#_Toc181020529)

[1.2: Ethical Considerations for Data Collection 3](#_Toc181020530)

[1.3: Chatbot Flow for Itinerary Planning in Dialogflow CX 4](#_Toc181020531)

[1.4: Chatbot Flow Demonstration 4](#_Toc181020532)

[Question 2: Evaluating Cloud Based Technology 4](#_Toc181020533)

[Question 3: Using Large Language Models 9](#_Toc181020534)

[3.1: Approach 9](#_Toc181020535)

[3.2: Analysis of Results 11](#_Toc181020536)

[3.3: Zero-Shot Classification for Boolean Validation 11](#_Toc181020537)

[3.4: Critical Analysis and Recommendations 11](#_Toc181020538)

[Question 4: Using Azure Machine Learning 12](#_Toc181020539)

[4.1: Insights 12](#_Toc181020540)

[4.2: Machine learning model 13](#_Toc181020541)

[4.3: Reflection on Using AzureML 14](#_Toc181020542)

[References 16](#_Toc181020543)

[Appendix 18](#_Toc181020544)

# Question 1: Build a Chatbot

## 1.1: Five Functionalities Related to Tourism for the Chatbot

**1. Attraction Recommendations**: The chatbot can recommend different places of interest depending on the user’s preference whether it is culture, beach or an adrenaline rush. This feature comprises ticket update and period of high demand for tickets for the users’ benefit (Sung & Kang, 2024).

**2. Itinerary Planning**: This chatbot can guide tourists on through the course of the day and what they can do since it can plan based on interests, location, and duration of visit. It also makes plans according to the occurrence of unpredictable events including weather information.

**3. Language Translation**: As for helping international guests, the chatbot provides the popular translations that will make it easier to communicate with the citizens of the country (Sung & Kang, 2024).

**4. Emergency Assistance**: In emergency situations, the chatbot suggests the location of the nearest hospital, police station or consulate so a person would be linked to assistance.

**5. Dining and Shopping Suggestions**: In line with the user’s choice and dietary preference, the chatbot reassures proposed eating establishments and convenient stores to visit around and offers any promotions (Sah, 2023).

## 1.2: Ethical Considerations for Data Collection

**Privacy and Consent**: This means that user permission to collect data must be sought before data collection is done. There are or should be the possibility to inform the users about which data is collected, how they will be used, and that the user can refuse from sharing such data but without prejudice to the mere purpose of the chatbot – answering questions (Sah, 2023).

**Transparency**: The types of data which are collected should also be stated and the intended use of data (for example, to provide a customized recommendation).

**Data Minimization**: Gather information that will be useful to monitor and assess their performance and do not collect other information. Restriction to the extent that only necessary information such as the users’ preferences is collected makes the data gathered not too much, nor too little and poses less security threats (Sah, 2023).

**Data Retention and Deletion**: Policies for data retention and procedures which should allow users to delete personal data. It fosters trust relationships and smooth working as it meets the data protection laws’ requirements.

**Data Security**: A user perspective would involve the provision of methods used in encryption and securing the users’ data in order to minimize on server side and external threats making user confidence in data safety high (Alla, 2024).

## 1.3: Chatbot Flow for Itinerary Planning in Dialogflow CX

In Dialogflow CX, the itinerary planning process of the chatbot simply initial by inviting the user and asking necessary information like travel dates, destination preference like Sydney, beaches etc or type of activities the user likes (Alla, 2024). For detail, constant entity picks in the field location, activity type, and duration are identified by Dialogflow’s NLP.

In addition, using these inputs, it provides day by day travel plan with recommendation about where to go next. This will include the use of APIs for acquiring the data of weather, timing and events making the plan more valuable (Alla, 2024). For flexibility, the chatbot options let the user change an activity or the time of an activity allowing the bot to re-estimate the options and then approve those changes.

## 1.4: Chatbot Flow Demonstration

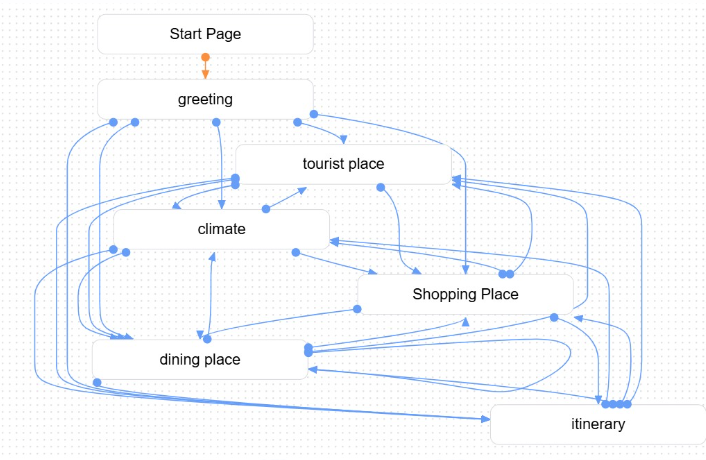


Figure 1: Chatbot flow

This chatbot flow begins with a greeting, and then provides options for tourists to explore specific areas: visitors’ attractions, weather, food and beverages, purchase, and time schedule. All the options lead back to the main menu, thus users can move freely between the features offered, giving the tourists an engaging and unique experience if they are looking for information on any aspect of their tour (Alla, 2024).

# Question 2: Evaluating Cloud Based Technology

**1. Content Integration**

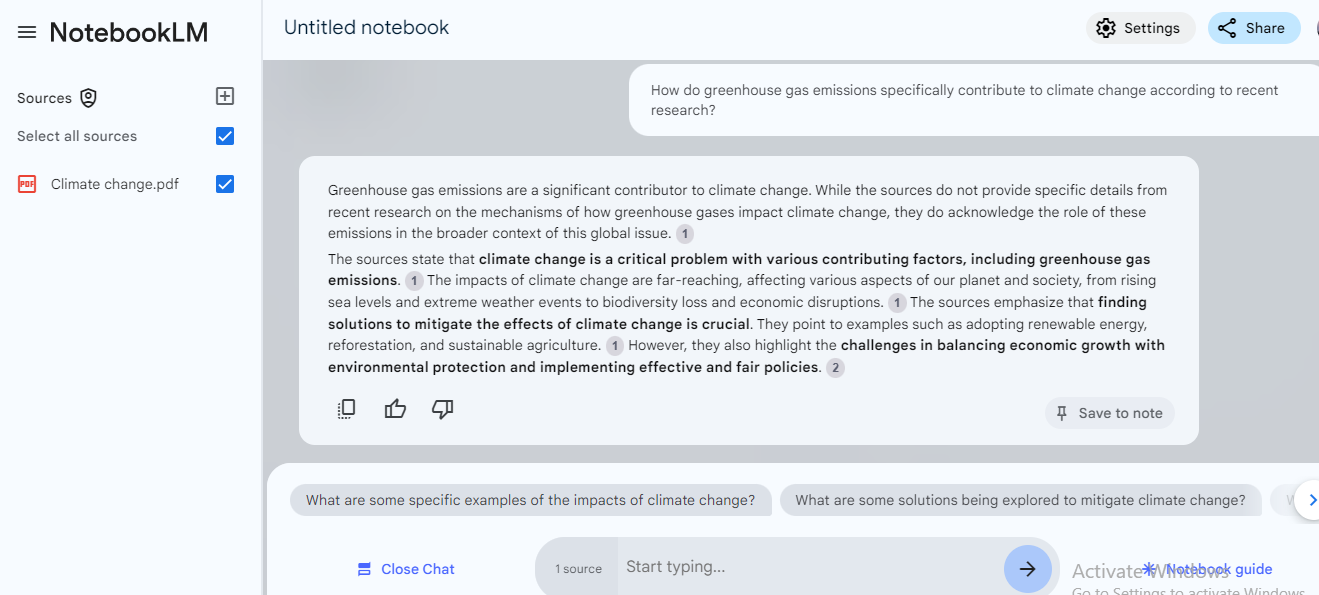
****

Figure 2: Content Integration

In NotebookLM, users can incorporate course content, research writings, textbooks, as well as other relevant information into the model. It also furnishes the student with sufficient context background that allows the model to reply to questions, condense the data, recognize associated concepts interrelated from different sources and interconnect them (Huffman & Hutson, 2024). They enhance the student ability to learn complex concepts by providing a single view of information sources. In a university context, this functionality makes learning faster by cutting down the amount of time students take to search for useful information and provides a much broader perspective on any academic subject.

**2. Real-Time Collaboration**

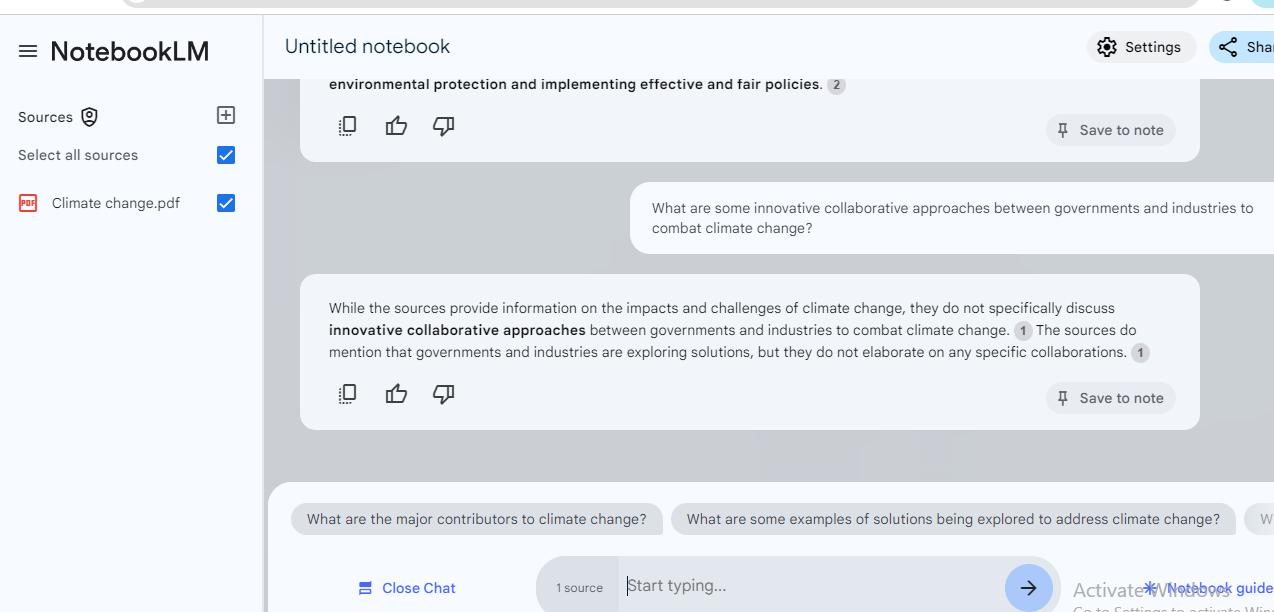
****

Figure 3: Real-Time Collaboration

Synchronous activities in NotebookLM involves various students, faculty or working on a project and working on the same NotebookLM document at the same time. It is most beneficial in group projects, research meetings and idea generating sessions since members can edit the document and add comments as well as notes to one another at the same time (Huffman & Hutson, 2024). AC 3 Collaboration enhances active participation since the students can instantly address what the other students are contributing thereby creating a more proactive learning process here fostered by technology.

**3. Enhanced Insights**

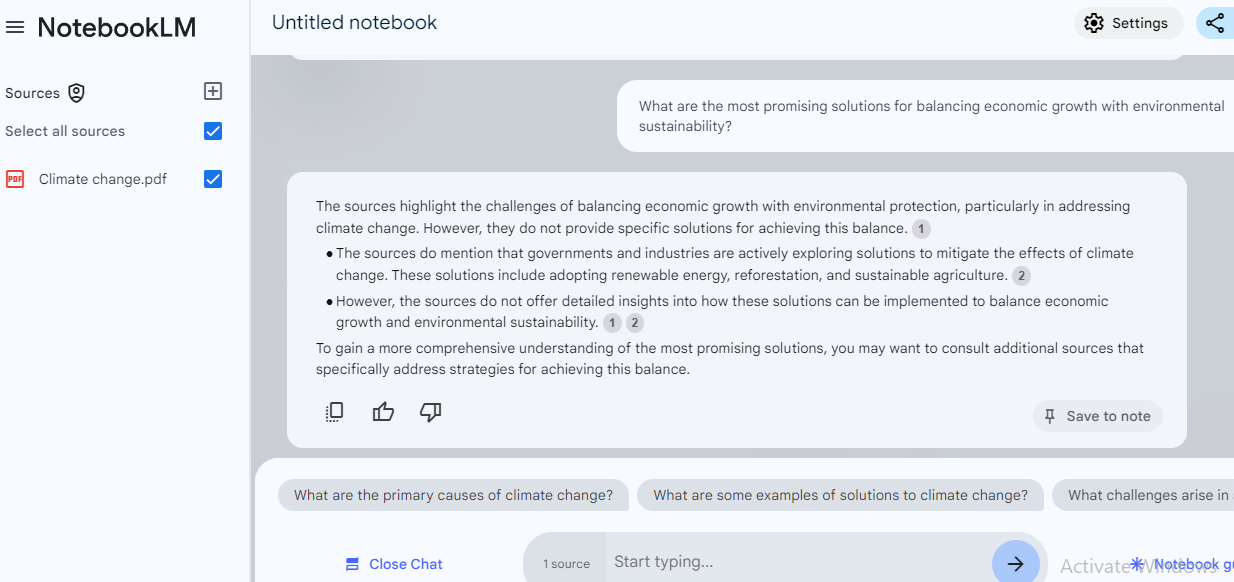
****

Figure 4: Enhanced Insights

Besides, the Enhanced Insights feature uses the language processing algorithm to give the learner more detail about a subject to make learning easier. Thus, using machine learning it is possible, for example, parse through what is uploaded by the users, identify key themes, and provide summaries and suggestions on what is worthy of further discussion (Kambhamettu et al., 2024). Penn Foster eagle is beneficial in learning as it assists the students get a broader view on the difficult concepts by asking the model to provide elaborative information, related information or the current information. It is even more helpful in areas such as research as well as on thesis development whereby the more information regarding the subject matter is useful.

**4. Question Generation**

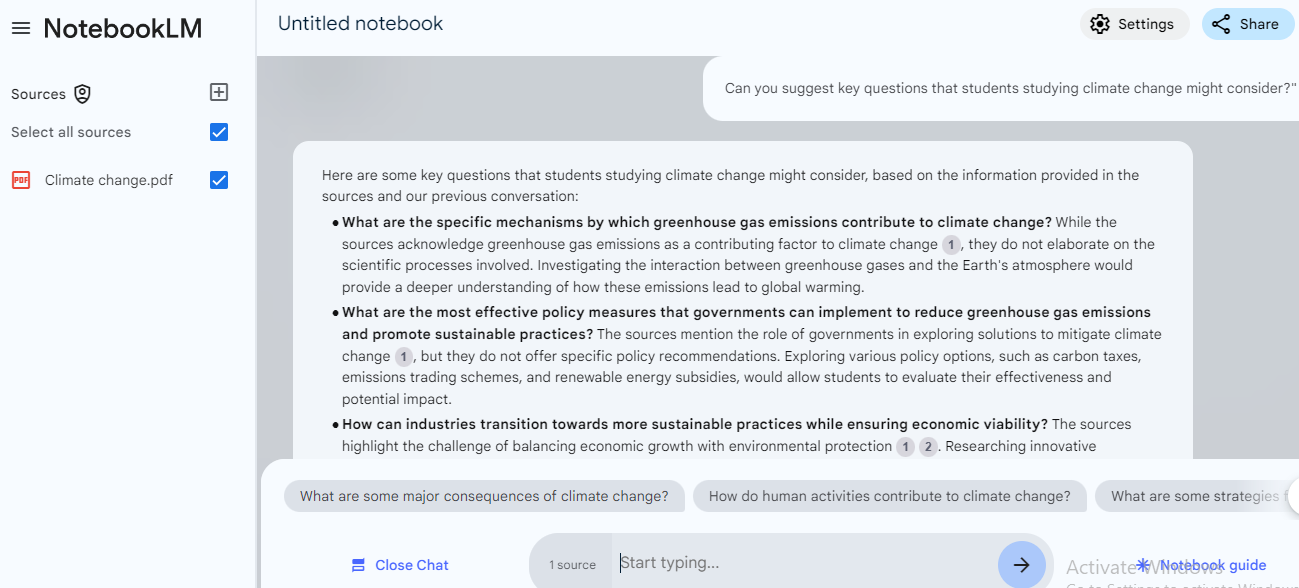
****

Figure 5: Question generation

The Question Generation function of NotebookLM supports learning since it generates valuable questions from the given content. This functionality helps students in self-quiz that pressure them to go through what they are reading, and even think of some crucial points on their own. It is also useful to the teachers, as the generated questions can be used for the preparation of tests, discussion sessions or a study session (Kambhamettu et al., 2024). Essentially, these questions direct students to areas of interest or lack of information hence may be used to supplement clarifications thus improving knowledge absorption on the course material.

**5. Feedback and Suggestions**

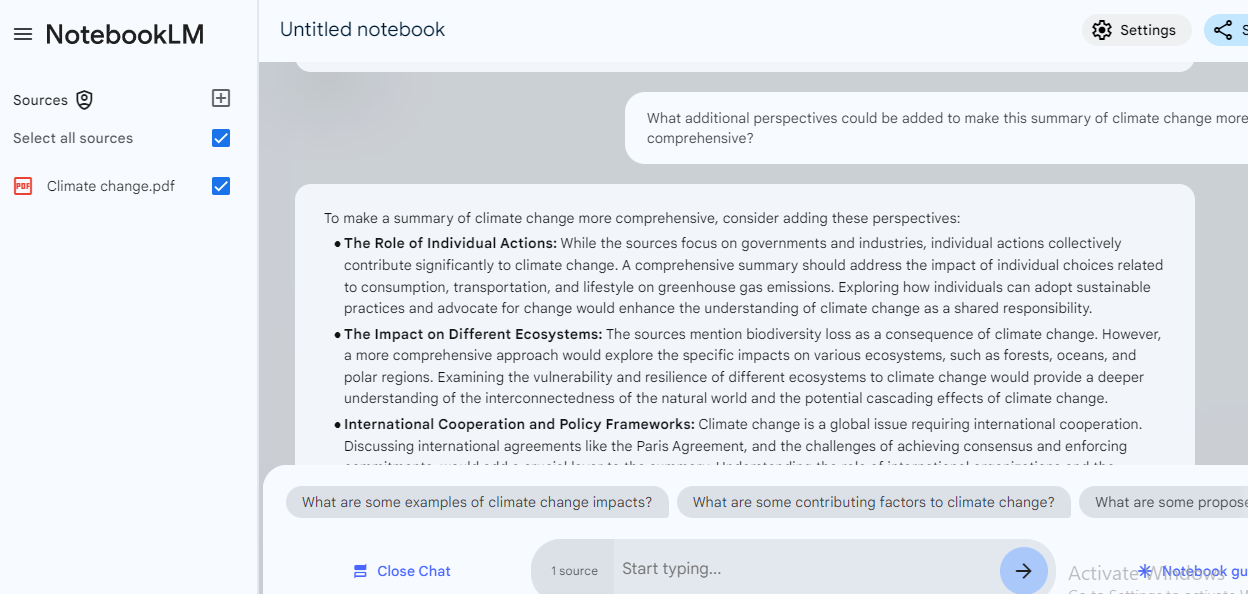
****

Figure 6: Feedback and suggestions

Users of NotebookLM can request constructive critique on their work or notes through the Feedback and Suggestions tool. This function can offer suggestions concerning more points to cover or other views and approaches or suggestions on making the content more intelligible (Tozuka et al., 2024). In a university, this feature can help the student in writing essays, reports, and presentation by providing additional refinements, missing viewpoints or important concepts. In this way, it motivates the process of learning, as well as contributes to the students’ producing broader and academically solid works.

# Question 3: Using Large Language Models

## 3.1: Approach

**Template: Number of Prompts Reviewed and Objective**

The template is to attempt to use follow up customer support based on customer reviews and employing an LLM to filter out specific entities like product name or product elements from reviews.

In the case of each extracted entity, the model should identify its sentiment as either “POSITIVE”, “NEUTRAL”, or “NEGATIVE” further, decide if there is a need to follow up, and if there is, then provide a reason for the following up (Tozuka et al., 2024). This process is in line with the overall mission to ensure all the dissatisfied customers get appropriate services at the right time.

**Example of the Prompt and Choices of the Sample**

When using an LLM such as GPT-3 from OpenAI or indeed, any model available via Hugging Face, you get a LLM that is given a prompt template which describes the expected JSON format output. This prompt template makes the model to come up with a JSON object that will have fields such as entity\_name, entity\_type, entity\_sentiment, followup, and followup\_reason (Chen et al., 2021).

**Code Implementation:** The code entails the creation of a zero-shot classification pipeline for the classification of sentiment for the entities as well as the individual overall sentiment of the review. The code itself features the analysis of text using manually-defined entities to consider the sentiment of the overall review text and decide whether follow-up is necessary**.**



Figure 7: Output for Question 3

## 3.2: Analysis of Results

Every entity is fine to pinpoint and the sentiment is correctly classified and the code also determines if a reply is necessary given negative sentiment in the review. Given the output analysis performed in this case, all the entities regiment a negative sentiment, meaning that the client had a bad encounter with the product and customer service (Chen et al., 2021). This creates a follow up activity for every client with a view to addressing complaints from its customers.

## 3.3: Zero-Shot Classification for Boolean Validation

To get an idea of whether entities generated represent accurate sentiments the sentiment classifications can be crosschecked with the help of a zero-shot classification model. This action is important to validate the LLM based analysis to existing models to support the sentiment scores assigned to each entity (Hadi et al., 2024).

## 3.4: Critical Analysis and Recommendations

**Accuracy**

The need to capture sentiment is particularly important in this model. The distinction between checking the individual entity sentiment helps to avoid general assumptions based on the overall sentiment of the reviews; this way, we make further relevant follow-ups when some components (e.g., product, service) require it (Hadi et al., 2024).

**Limitations**

While helpful, LLMs fail to interpret domain-specific words particularly well unless prompted as would bring about relevant details and subtle differences. Moreover, entities which were extracted manually in this example could be also automated using NER models for retail specific (Hadi et al., 2024).

**Recommendation**

Continue the machine learning process using an online specialty dataset to label sentiment and other follow up criteria. This will enhance the results especially when one is dealing with the cross over reviews that are either positive or negative.

# Question 4: Using Azure Machine Learning

## 4.1: Insights

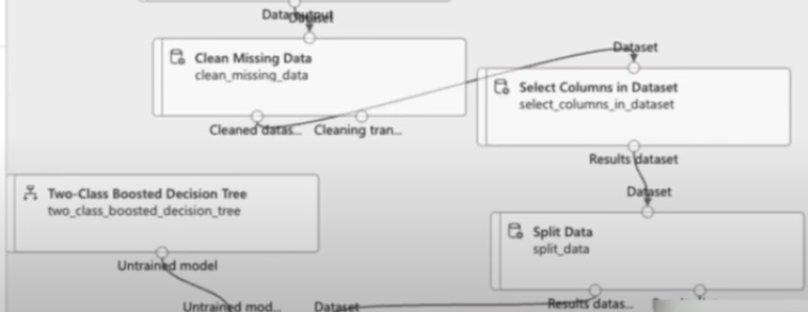


Figure 8: Implementing models

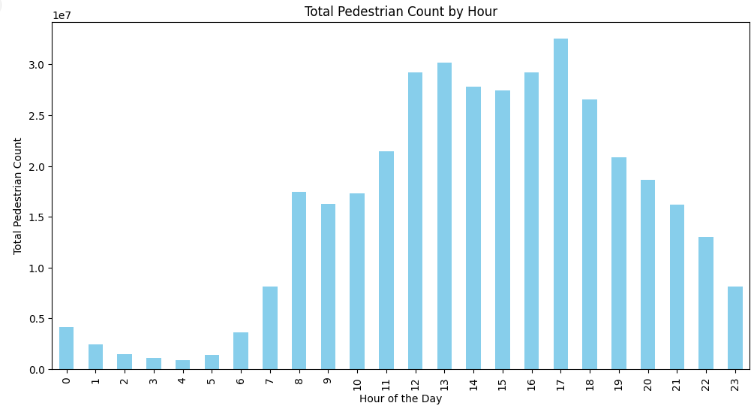


Figure 9: Total pedestrian count by hour

This chart shows times at which pedestrian traffic is high and low using bar graphs to differentiate the hours of the day. Local higher counts during certain hours refer to crowded people during the morning and evening rush hour time (Imad Zeebaree, 2024). Such knowledge is useful in assigning priority to the city’s infrastructure and service delivery during high pedestrian traff

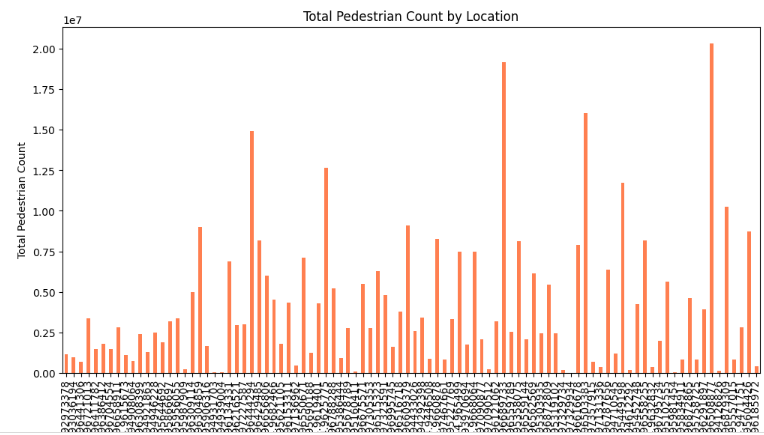


Figure 10: Total pedestrian count by location

The following chart shows the volume of pedestrians in different places at Melbourne CBD. Some areas demonstrate a higher traffic a constant value showcasing that some zones may be more frequented than others (Imad Zeebaree, 2024). It facilitates the planner to determine the zones to direct the resources and where improvements to the physical city environment should be made in order to improve the city navigation and walkability for pedestrians in the busiest regions.

## 4.2: Machine learning model

**1. Regression Analysis and random forest**

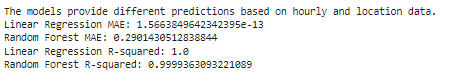
****

Figure 11: R-squared value

Here in the particular AzureML task, two models were considered for providing the best possible predicted count of the pedestrians. The outcome of the Linear regression model seems to be excellent the MAE value is equal to 1.57e-13/neary zero and R squared is 1.0— represents the perfect accuracy to explain the variability in the data. This implies that it is very appropriate for this dataset. Random Forest model is also quite stable with basic metric MAE of 0.29 and R-squared of 0.9999 that closes to perfect model as its claiming of explaining nearly databased variability with minor error to linear regression (Ng’ang’a et al., 2023). These results show that both these models are efficient, although linear regression performs marginally better given the data set’s linear characteristics and thus will be ideal for forecasting of pedestrians’ activity in Melbourne CBD.

## 4.3: Reflection on Using AzureML

**1. Data Preparation Pains**

There is no external data quality challenge that the author had during the data preparation when using AzureML but time and count features were an issue during the preprocessing. To this end, I routinely used the data pre-processing tools within AzureML to perform missing data imputation using the mean and formatting date-time variables properly (Ng’ang’a et al., 2023). Such adjustments were vital for models since it allowed the data to be structured, which is necessary for training.

**2. Model Creation And Evaluation**

The second difficulty has been outlined with the model selection and evaluation in AutoML. AzureML comes equipped with an AutoML function which somehow gives multiple models that has different performances such as R-squared value and mean absolute error. In order to choose the best model, it was crucial to analyze performances of each model in terms of evaluating the accuracy with the possibility to also interpret the results (Ng’ang’a et al., 2023). When the AutoML experiment had been completed I compared each model and settled for one with the highest R-squared and the lowest error rate for the task, to yield the most accurate prediction of pedestrian counts.

**3. Overcoming Technical Constraints**

Utilizing AzureML’s environment was not straightforward due to some challenges that were realized while setting up parameters of AutoML. However, thanks to Azure’s documentation and guides, I was able to effectively work with these tools and get the best fit for the prediction of pedestrians (Ng’ang’a et al., 2023).

# References

Alla, E. (2024). Web-Based IFC Application for Infrastructure Models: Enhancing User Interaction Through Natural Language Processing - Webthesis. *Polito.it*. https://webthesis.biblio.polito.it/secure/32498/1/tesi.pdf

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., & Ryder, N. (2021). Evaluating Large Language Models Trained on Code. *ArXiv:2107.03374 [Cs]*. https://arxiv.org/abs/2107.03374

Hadi, M. U., Tashi, Q. A., Shah, A., Qureshi, R., Muneer, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., Mirjalili, S., & Shah, M. (2024). *Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects*. https://doi.org/10.36227/techrxiv.23589741.v6

Huffman, P., & Hutson, J. (2024). *Enhancing History Education with Google NotebookLM: Case Study of Mary Easton Sibley’s Diary for Multimedia Content and Podcast Creation*. Digital Commons@Lindenwood University. https://digitalcommons.lindenwood.edu/faculty-research-papers/683/

Imad Zeebaree. (2024). The Distributed Machine Learning in Cloud Computing and Web Technology: A Review of Scalability and Efficiency. *Journal of Information Technology and Informatics*, *3*(1). https://qabasjournals.com/index.php/jiti/article/view/281

Kambhamettu, H., Flores, J., & Head, A. (2024). *Traceable Text: Deepening Reading of AI-Generated Summaries with Phrase-Level Provenance Links*. ArXiv.org. https://arxiv.org/abs/2409.13099

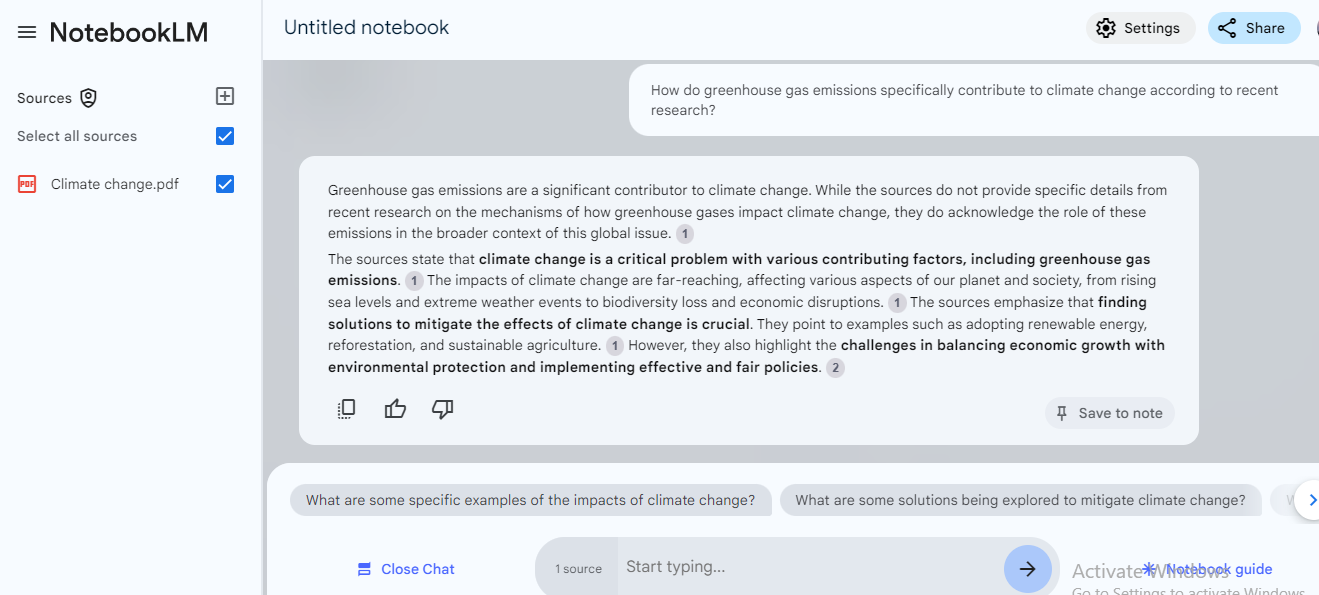
Ng’ang’a, D. N., Cheruiyot, W. K., & Njagi, D. (2023). A Machine Learning Framework for Predicting Failures in Cloud Data Centers -A Case of Google Cluster -Azure Clouds and Alibaba Clouds. *Research Square (Research Square)*. https://doi.org/10.2139/ssrn.4404569

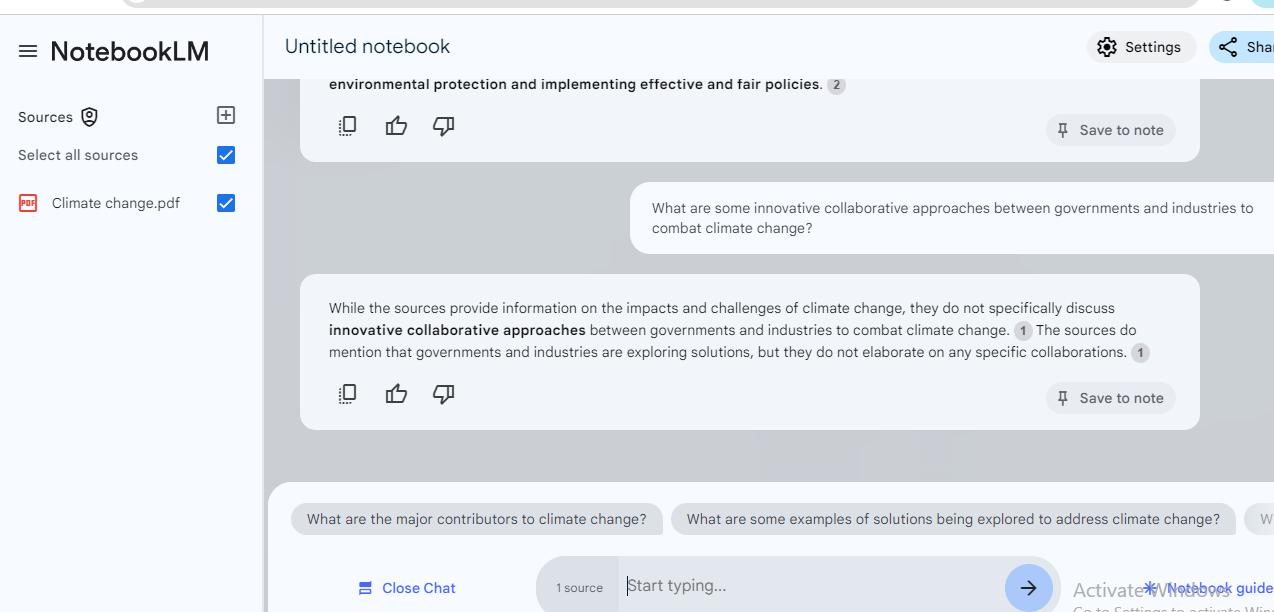
Sah, A. (2023). Building a chatbot using dialogflow : a proof-of-concept demonstration. *Lutpub.lut.fi*. https://lutpub.lut.fi/handle/10024/166650

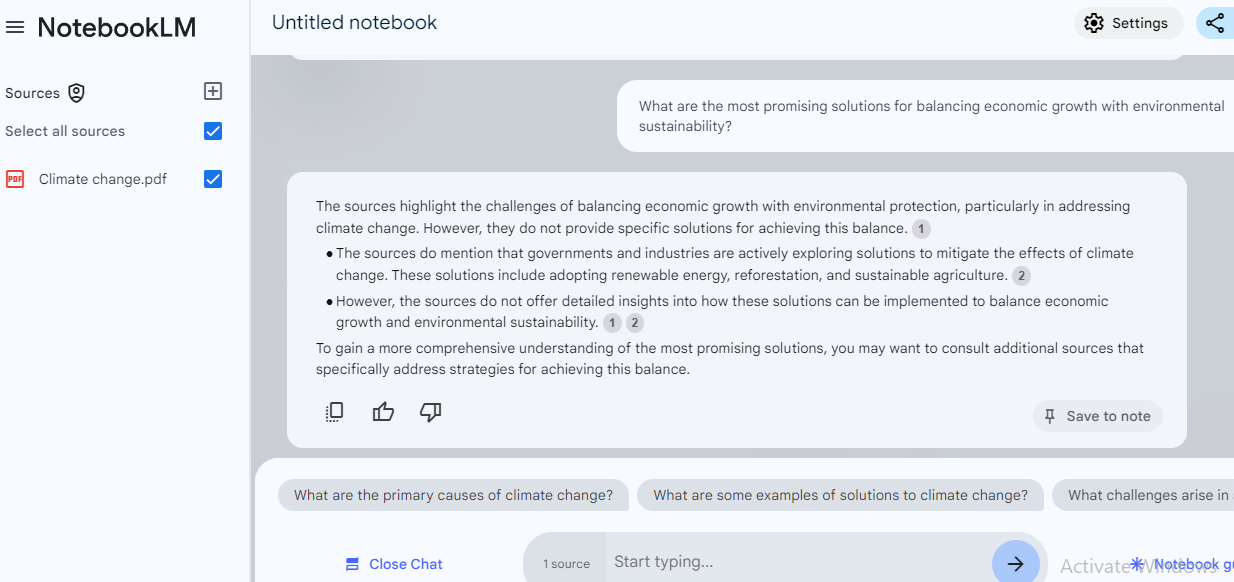
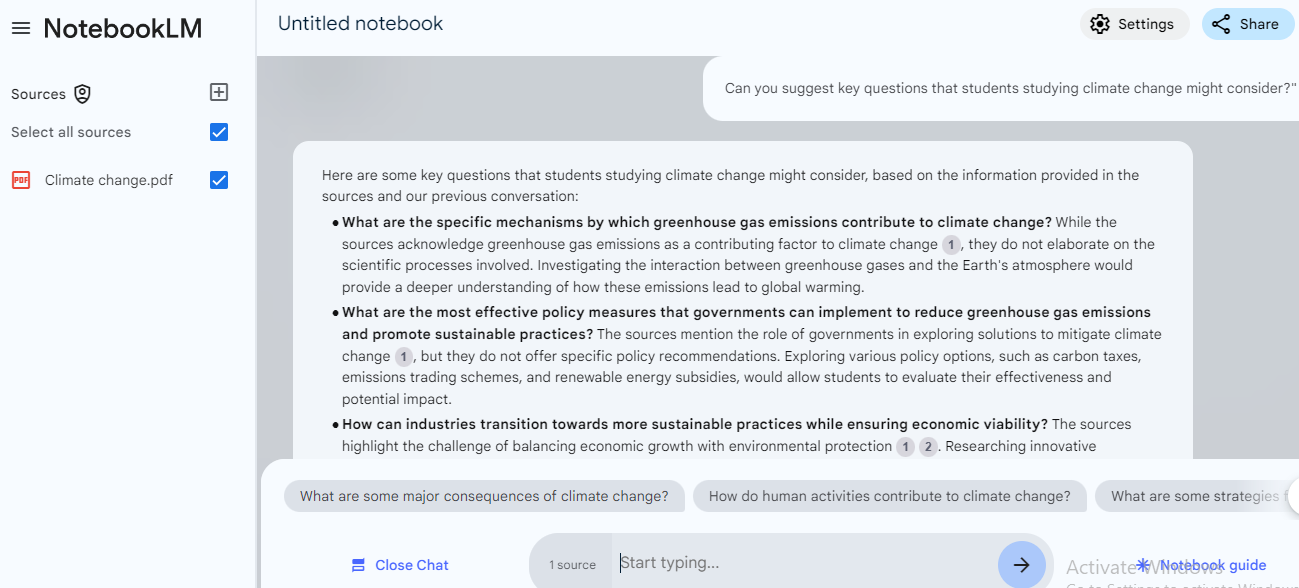
Sung, M.-C., & Kang, S. (2024). Developing AI Chatbots for Pragmatic Instruction of Korean Secondary L2 English Learners. *Yeongeo Hak/Yeong’eohag*, *24*, 441–459. https://doi.org/10.15738/kjell.24..202405.441

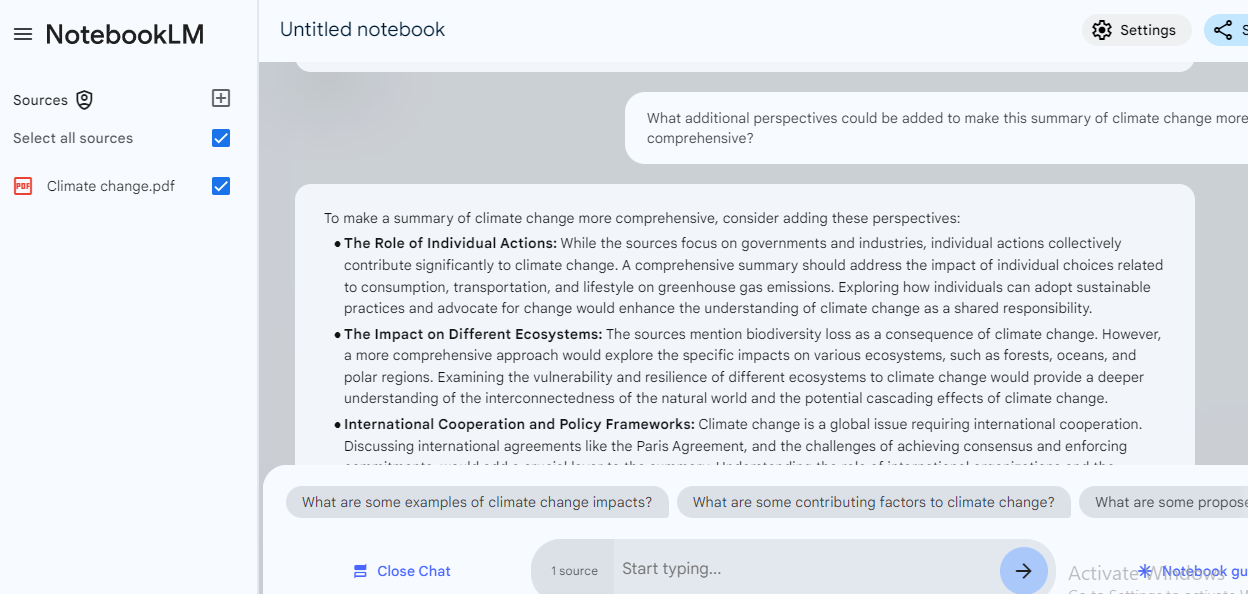
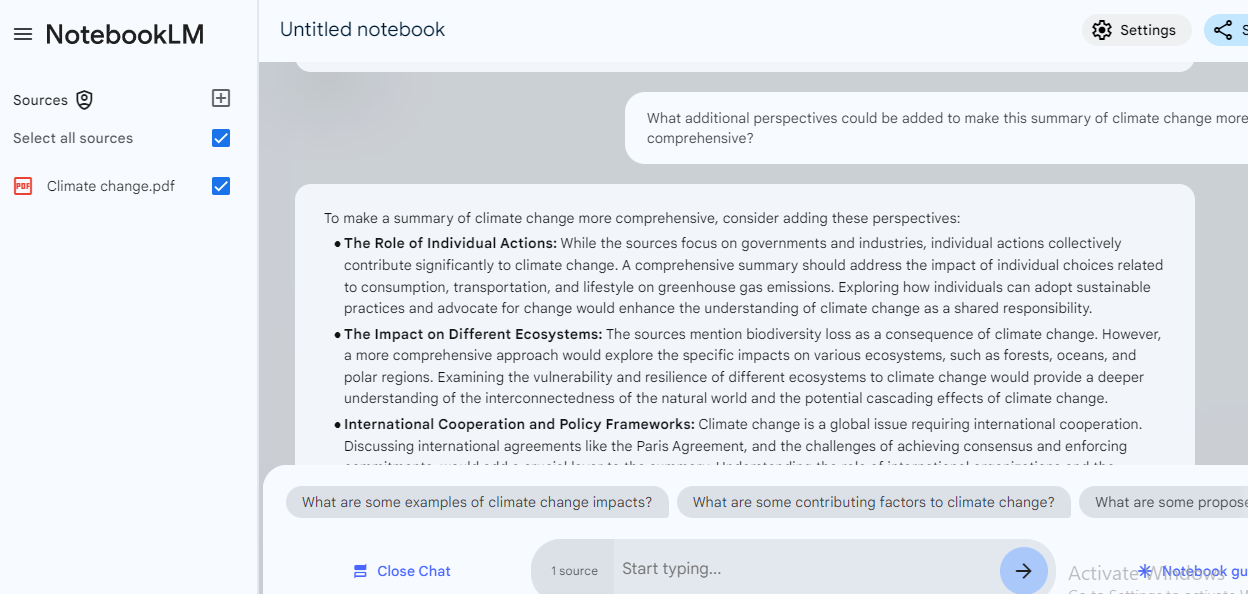
Tozuka, R., Johno, H., Amakawa, A., Sato, J., Muto, M., Seki, S., Komaba, A., & Onishi, H. (2024). *Application of NotebookLM, a Large Language Model with Retrieval-Augmented Generation, for Lung Cancer Staging*. ArXiv.org. https://arxiv.org/abs/2410.10869

# Appendix





’

****’

****

