Designing a Database for Cambodia's Foreign Direct Investment and Trade

Chakrayuddh Kimsovanna

EPPS 6354 – Information Management

Dr. Karl Ho

University of Texas at Dallas

May 7, 2025

Table of Contents

Introduction

Foreign direct investment (FDI) and international trade are two of the most important indicators to consider when analyzing how well a country's economy is performing. Cambodia, being a small nation in Southeast Asia, has been rapidly developing over the last few decades, as it continues to recover from years of conflict and political instability during the Cold-War and its aftermath. As it opens itself more to global markets, Cambodia has attracted many foreign investors thanks to its abundance of natural resources, as well as labor. With a huge inflow of FDI and new bilateral trade agreements with both regional and global partners, many sectors of its economy are booming, particularly in manufacturing, real estate, tourism and construction (World Bank Group 2024). Cambodia's GDP is about $42.34 billion dollars, and according to the International Monetary Fund, its real GDP is projected to grow from 5.5 percent in 2024 to 5.8 percent in 2025 (IMF 2025). These developments have contributed to the creation of hundreds of thousands of jobs and an overall improvement in the standard of living despite the relatively high levels of people still living in poverty.

As a Cambodian myself and a student with a background in economics, I am especially interested in understanding how my country's economy functions. However, existing data on this topic is often scarce and difficult to access, especially from the Cambodian government sources which typically release only lengthy reports, sometimes written in Khmer rather than English. Visualizations are also limited in this regard. While some international organizations like the World Bank do provide trade data on Cambodia, they do not integrate it with any information regarding the FDI. To analyze this effectively, it is essential to have more accessible and well-organized data. That is why the purpose of this project is to not only examine the relationship between FDI and international trade, but also to improve data accessibility and user convenience

through the development of a database application. This tool is not necessarily intended to serve as a definitive or official source of data, but rather as a prototype of what a more comprehensive and also interactive system could look like, so that future researchers, economists or even policymakers could use it to examine Cambodia's investment and trade patterns more effectively.

Methodology

Data Collection

The primary data sources I used were Open Development Cambodia (ODC) for FDI information and the World Integrated Trade Solution (WITS) for trade data. ODC is a non-governmental, non-profit organization that collects and compiles data from official sources, like the Council for the Development of Cambodia (CDC)—a government agency involved specifically in overseeing investment activities—as well as news articles and company websites. As for the WITS, it is a site that is managed by the World Bank that provides all the information related to trade, mostly compiled from the UN Comtrade database. However, the data here is not as up-to-date as I wanted it to be, as the latest available data from WITS is 2022, while ODC last updated its website back in 2023. One reason that might explain this delay is that especially for WITS that relies on UN Comtrade—which in turn relies on trade statistics submitted by each individual country—there could be extensive verification process before new data is published. The only other sources of information about Cambodia's exports and imports that are slightly more recent (2023) were the Observatory of Economic Complexity (OEC) and TrendEconomy; however, the former is locked behind paywall and the latter's data source is unreliable.

The types of economic sectors explored in this project include industrial development (i.e., manufacturing), commercial real estate, energy, agriculture, infrastructure, finance, tourism and mining. For the product categories listed in the trade data, I follow WITS's use of the HS

1988/92 classification created by the World Custom Organization (WCO), which is one of the standards for grouping the types of products; these include animal, vegetable, food products, minerals, fuels, chemicals, plastic or rubber, hides and skins, wood, textiles and clothing, footwear, stone and glass, metals, machinery and electronics, transportation, and other miscellaneous goods.

For the scope of this project, I focused only on the top 11 investor economies based on their number of FDI projects in Cambodia. These include China, Japan, Vietnam, South Korea, Singapore, France, the United States, Thailand, Malaysia, as well as Hong Kong and Taiwan. Even though Hong Kong is a part of China with its special administrative status, it is still listed separately from China in the datasets, so it is justified to have Hong Kong as its own entity. As for Taiwan, while it is a de facto sovereign state with its own independent government, it is not a member of the UN and so data about its trade with Cambodia was not in the WITS database, which means it will be excluded from the import-export analysis.

More than 400 different investment projects are listed in the ODC database, but in this project, I selected only 225 of them, as I excluded all projects that are not developed by any of the aforementioned countries, as well as those that involved Cambodian-owned firms as co-investors in order to maintain consistency when looking at foreign-led investments. For example, the ongoing Funan Techo Canal project, while partially funded by a Chinese-owned company, is still mostly owned (51 percent of the share) by Cambodian firms (Pham 2024). Another issue I ran into was that many companies either had no reported country of origin or were listed as Cambodian-owned, even though their parent companies are foreign, usually Chinese. To address this, I manually checked each of these company and tried to verify its origin through all the

available online sources. For those that were too ambiguous or simply unverifiable, they were excluded from this data.

Data Processing and Schema Design

To organize all this data, I first created Excel files to easily review the information I gathered from ODC and WITS. While the ODC does provide `.csv` files for its dataset, there were many unneeded variables and I noticed some inconsistencies in how they report the companies' names—for example, there are companies whose names ended with "Co., Ltd," while some ended with "Company Limited." In another case, there is one company labeled as "Huadian Sihanoukville Power Generation Co., Ltd" and another called "China Huadian Corp Ltd" even though both are the exact same entities with the same directors (ODC 2023). One crucial fact to highlight is that there are investment projects that are co-invested by multiple companies, sometimes from different countries (e.g., Japan and France both invested in a project in Battambang province to build a drainage system), and some even spanning over multiple provinces, while there are firms that invest in multiple projects in the same province. The complexity here shows the need for my relational database to also include many-to-many relationships, particularly between projects, investor countries and provinces.

I used pgAdmin 4 to populate the data to build my relational database. As shown in the entity-relationship diagram that I designed below (Figure 1), this database consists of a total of ten tables. The main table is `fdi_project` which shows all the development projects used in my research. Its primary key is `fdi_id` and it includes key variables `project` (name of the project), `sector` (economic sector), `investment` (amount of money invested, in millions of USD), and `job_creation` (estimated number of jobs created). To reflect the many-to-many relationships as mentioned earlier and solve any normalization issue, I had to create new linking

tables: `project_country`, `project_company` and `project_province`, where I assigned a unique ID to every investor country, company and province, and then associate them with their respective `fdi_id`. In the table `company`, I simply assigned each company name to its ID; the same process goes with the `province` and `investor_country` tables. These IDs were then used to link the entities to the corresponding FDI projects through the three many-to-many linking tables described above.

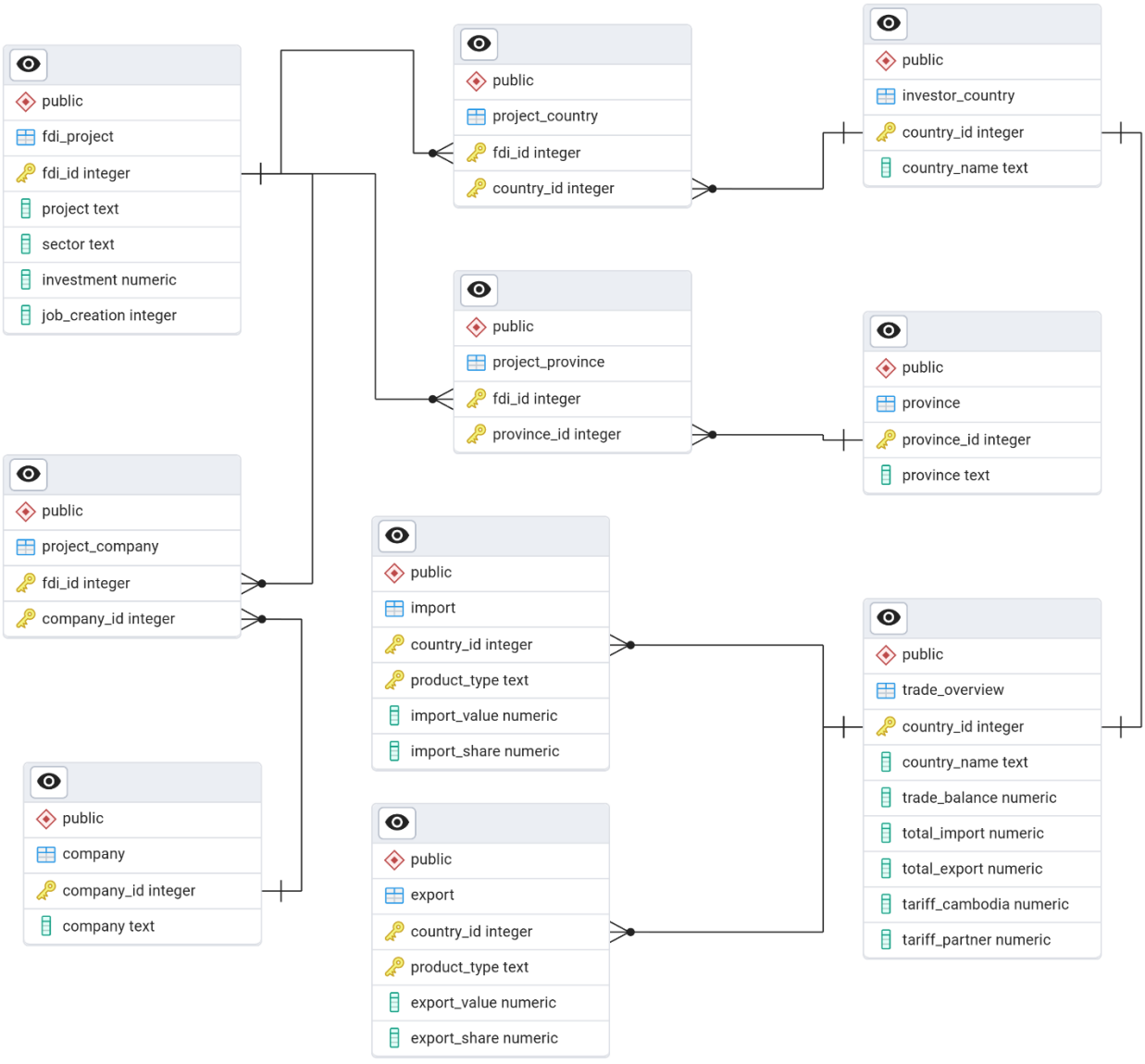

Figure 1. Entity-relationship diagram of the database schema.

The next main table is `trade_overview`, which incorporates the trade data into the database. This includes variables, such as `country_name` (name of the investor country), `total_import` (total import values in millions USD), `total_export` (total export values in millions USD), `trade_balance` (calculated as total exports minus total imports). Two additional variables are tariff rates: `tariff_cambodia` (AHS-weighted average or the average tariff imposed by Cambodia on foreign goods) and `tariff_partner` (AHS-weighted average or the average tariff imposed by the investor country on Cambodian exports). Connected to this table are two dependent tables—`import` and `export`—which are weak entities because they each rely on the `country_id` in `trade_overview`. Each of these tables contains more detailed data, like `product_type` (type of goods as defined by the HS 1988/92), its respective `import/export_value` as well as `import/export_share`.

I would like to note that due to time constraint, I had to use an AI tool, Copilot, to help me convert the information I gathered in the Excel file for FDI projects into SQL code, as I had already designed how each table was supposed to interact with one another. As for the trade data, unfortunately, Copilot failed to maintain consistency for the numbers, so I was forced to manually insert the numbers myself.

Database Design and Deployment

After populating the data and designing the schema in pgAdmin 4, I used RStudio to implement the database and publish it online. I used the `DBI` package to allow me to send SQL queries to my database. However, I soon realized that doing any publishing would not work as the database was still locally hosted. To circumvent this issue, I converted the SQL schema that I wrote in pgAdmin 4 (saved as a `.sql` file) into a `.sqlite` version directly within RStudio via

the RSQLite package instead of using DB Browser. This allowed me to connect the database to the Shiny server successfully without any disconnection error when publishing.

Before I built the application itself, I had to properly write the SQL query to set the foundation first. As shown in Figure 2, to generate the main FDI project table, I used a series of `left join` operations to connect the three linking tables described earlier, linking each FDI project to its corresponding companies, investor countries and provinces. I then used Copilot to help me understand a very useful function, `group_concat`, which allowed me to combine multiple values into just a single string separated by a comma without having repeated rows. For imports and exports, as well as trade overview, the SQL queries were a lot simpler, as I only needed to select the variables from those tables.

```
query <- "
  SELECT f.sector, f.project, f.investment, f.job_creation,
    GROUP_CONCAT(DISTINCT c.company) AS companies,
    GROUP_CONCAT(DISTINCT t.country_name) AS countries,
    GROUP_CONCAT(DISTINCT p.province) AS provinces
  FROM fdi_project f
  LEFT JOIN project_company pc ON f.fdi_id = pc.fdi_id
  LEFT JOIN company c ON pc.company_id = c.company_id
  LEFT JOIN project_country pt ON f.fdi_id = pt.fdi_id
  LEFT JOIN investor_country t ON pt.country_id = t.country_id
  LEFT JOIN project_province pp ON f.fdi_id = pp.fdi_id
  LEFT JOIN province p ON pp.province_id = p.province_id
  GROUP BY f.fdi_id, f.sector, f.project, f.investment, f.job_creation
"
```

Figure 2. SQL query for the main FDI project table.

The user interface, or UI, for this database application was made using Shiny's `fluidPage` layout, where I created a `titlePanel` that shows the title of my database and a `tabsetPanel` that consists of five subpanels, including "All FDI Projects," "Trade Overview," "Imports," "Export," and "Map of FDI Projects." All the tables displayed in the database were created using the `DT` package so that they can be interacted with (i.e., sorting by descending or ascending order) and are searchable. Because I wanted to have filters for more user control, I had

to rely on another AI tool, ChatGPT, which guided me on how to make these dropdown menus actually work. Simply having the `selectInput` functions in a tab panel was not enough, as I had to use the `observe` function put inside the server codes as well, which would allow me to pull certain values from the database and load them into each dropdown; here, I created filters for investor country, sector and province (shown in Figure 3).

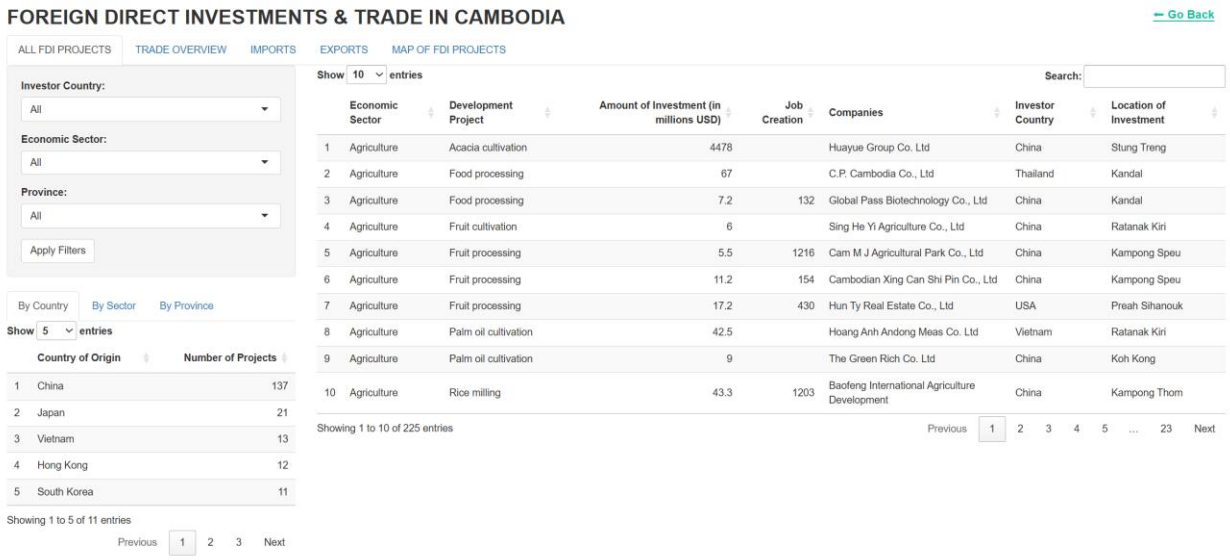

Figure 3. Screenshot of the "All FDI Projects" tab.

As seen in the bottom left corner of the page (Figure 3), there is a mini set of data tables that provides some statistics related to the total number of projects invested by a certain country, as well as within any specific sector and province. I simply created this by creating another `tabsetPanel` within the main `tabsetPanel`. As with the rest of the UI layout, I adjusted the positions and sizes of each element through trial and error until I managed to get the application to look relatively clean and easy to navigate. Furthermore, in the Import and Export tabs, I decided to include some visualizations in addition to just their tables. To create the bar charts and the pie charts, I used the `ggplot2` package to plot trade values and share, respectively, by product type as I used a query to connect to either the `import` or `export`

tables. However, I wanted to make it change dynamically for each country that I would select in the dropdown menu, so with ChatGPT's help, I had to insert this code when rendering the plots: `req(input$import_country)` (or `req(input$import_country)`) followed by `params = list(input$import_country)`, both of which wrap around the SQL query that fetches the data (Figure 4). Once again, I had to keep tweaking the UI until the layout looked balanced and everything was properly aligned without awkward positioning.

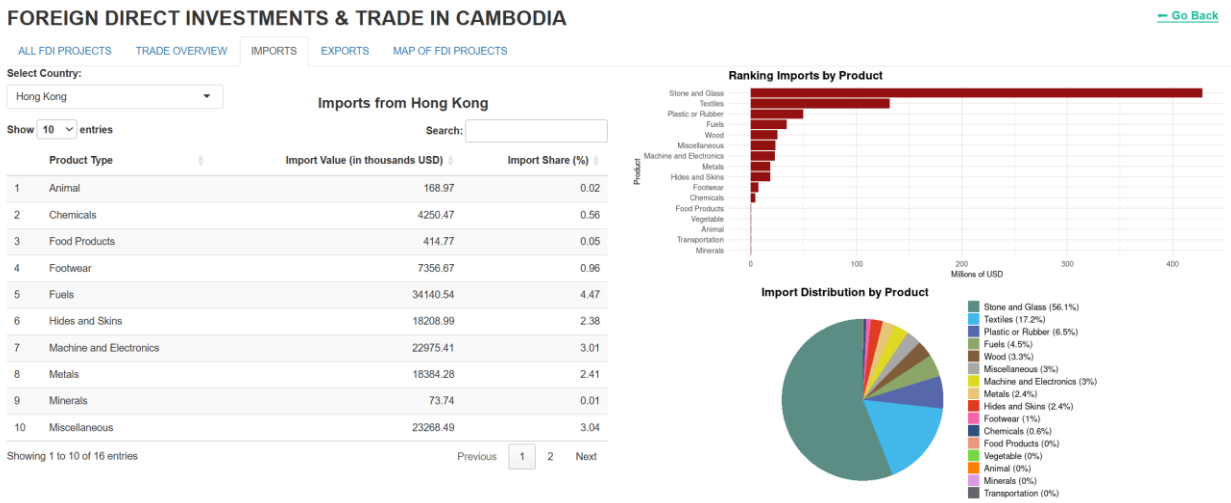

Figure 4. Screenshot of the "Imports" tab, showing data about imports from Hong Kong to Cambodia.

The final tab, "Map of FDI Projects," was by far the most difficult part to implement due to its complexity of dealing with geospatial data as well as having interactive visualizations. Again, as a disclaimer, I relied on both ChatGPT and Copilot to guide me step-by-step through this process. I did not copy and paste entire code blocks; instead, I used these tools to generate basic code structures, which I then modified, from inserting SQL queries to renaming variables to adjusting various visual elements. However, I did resubmit these codes I had modified back to Copilot to check for any missing brackets or syntax errors, especially when I encountered issues while running the application. First, I downloaded a `.geojson` file of the map of Cambodia

that contains all of its administrative boundaries from the GeoBoundaries dataset, which is sponsored by the UN Office for the Coordination of Humanitarian Affairs (OCHA) (GeoBoundaries 2023). Due to some mismatches between how each province's name was labeled in that `.geojson` file, I had to rename several of them to maintain consistency.

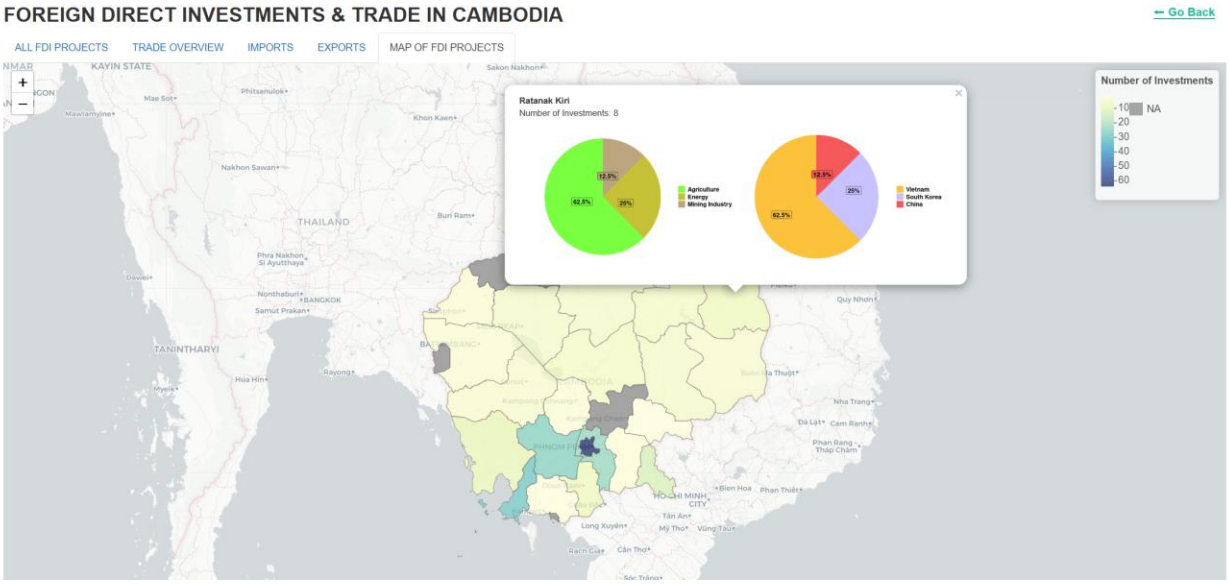

Figure 5. Screenshot of the "Map of FDI Projects" tab, showing the number of investment projects and their distribution by investor countries and sectors in Ratanakiri province.

Afterwards, I used the `leaflet` package to render an interactive map where each province is shaded based on the number of FDI projects it has. I wrote an SQL query that counted the number of projects per province and then used it to match each province by name with its respective boundary in the map. Next, I created pie charts using `ggplot2` to show the distribution of projects invested by investor country and by sector. I initially tried showing the percentages using just the `geom_text` function from `ggplot2` but the labels quickly became cluttered, making it almost impossible to read the pie charts that have smaller slices. With ChatGPT's suggestion, I installed the `ggrepel` package so now I could make these percentage labels a little easier to see as they now can adjust their positions automatically, which is useful

for these small slices. While it is not perfect, the result is much more readable. I then assigned custom hex color codes for each sector and country, so that they retain the same color across different popups. To display these charts in the popup for each province, I was again guided by ChatGPT to convert them to base64-encoded PNG images using the packages `ggsave` and `base64enc`, which would allow these charts to be embedded directly into the popup's HTML. I could have skipped the base64 part and just used `ggsave` to save each chart as image files, then uploaded and embedded them individually into the popup content, but this helpful method made the whole process much less tedious. The end result is as shown in Figure 5, where one could see the total number of investments the province of Ratanakiri has, along with two pie charts showing the share of the project types and the countries that invested in them. Overall, this interactive map serves as the main highlight of this entire application, providing viewers with a clear visual summary of investment patterns across different regions of the country.

Finally, with the database application successful uploaded and running on the Shiny app, I embedded its HTML into a Quarto document, where it would be rendered on RStudio and uploaded into my Github repository, so now it becomes viewable directly from my Github website.

Key Findings

Using this application, I was able to examine Cambodia's FDI and trade patterns. Of the reported 225 different investment projects, 93 of them belong to the industrial development sector, followed by commercial real estate (39 projects), energy (28 projects), agriculture (26 projects) and infrastructure (23 projects) (Kimsovanna 2025). The financial sector, along with tourism and mining industry, have less investment projects, together totaling up to only 16 projects. In terms of how many projects are invested by a foreign country, China easily tops the

group with 137 projects. The second place goes to Japan at 21 projects, followed by Vietnam with 13. The next four investors include Hong Kong, South Korea, Taiwan and Singapore. This trend is notable, as it shows that East Asian countries, or broadly speaking, countries within the Sinosphere (including Vietnam and Singapore), are the dominant investors in Cambodia. It is somewhat surprising that Western countries, such as France and the United States, show very low levels of investment relatively. Thailand and Malaysia are towards the bottom of the list, with just 4 projects each. In addition, the reason why industrial development is so prominent is because a lot of the investments come from clothing and shoes manufacturing; according to the Council for the Development of Cambodia's recent report, the country is "the world's eighth-largest exporter of apparel and footwear. Over one-third of the nation's GDP is generated by the textile and apparel industry" (CDC 2024).

The commercial real estate sector is booming, with new hotels and casinos, as well as apartment buildings and shopping malls, opening in major urban areas, such as Phnom Penh and Sihanoukville. Cambodia is also known for its rich, fertile lands, which explains why many foreign companies are buying or leasing large hectares of lands to produce rubber and crops, like rice and sugarcane—as shown in this database, both China and Vietnam are leading in this area (Kimsovanna 2025). From Phnom Penh, where the four main rivers meet, one can enter every province by just sailing along the interconnected tributaries and streams. This geographic feature makes it ideal to invest in many hydropower dams, which account for 19 out of the 28 energy projects. There are 6 solar farm projects, invested by a multitude of countries like the United States and Thailand, but they are much smaller in scale compared to hydropower energy projects. Infrastructure development is another major area of investment, with projects ranging from

roads, railways and bridges to drainage systems, hospitals and airports. Again, China is leading in this sector as well, followed by Japan.

In terms of development patterns across regions, Phnom Penh clearly has the highest number of investment projects, given its role as the capital and the country's center of trade and commerce. About half of those investments are in the commercial real estate sector (Kimsovanna 2025). Provinces with relatively high levels of investment include Kandal, Kampong Speu and Preah Sihanouk. This is expected, as Cambodia's main gateway to the world lies in the port city of Sihanoukville (located in the southwestern tip of the country), and thus, most investment projects would be located along the corridor connecting Preah Sihanouk province (also known as Kampong Saom) and Phnom Penh. Other provinces like Stung Treng and Koh Kong serve as key energy-producing regions, with several hydropower dams built along their rivers, while Preah Vihear, Kampong Thom and Kratie (pronounced "Kracheh") are more focused on agricultural developments.

Moreover, Cambodia's trade activity reveals some similar patterns to its FDI trends. As of 2022, the country's total amount of exports was valued at $20.58 billion, while its total imports were valued at $29.94 billion (WITS 2022). As shown in Figure 6, Cambodia mostly imports from China, reaching to a total of about $10.45 billion, while its exports to China is much lower at just $1.24 billion. On the flip side, Cambodia exports the most to the United States of almost $9 billion worth of goods, while it only imports less than a third of a million dollars from there (Kimsovanna 2025). As a result, Cambodia manages to achieve a massive trade surplus with the United States, but a huge deficit towards China, at $8.66 billion and negative $9.20 billion respectively. The only other countries that Cambodia has a trade surplus with are Japan and France, while it runs trade deficits with all the others. In terms of tariffs, Cambodia imposed high

rates on goods from Japan (15.53%), South Korea (13.49%), the United States (12.45%) and Thailand (11.07%). Other countries also face tariffs, but at more moderate levels. Only Singapore enjoys an average tariff rate of less than 1 percent. Likewise, Malaysia and the United States imposed their own tariffs upon Cambodian exports at 9.45 and 7.95 percent, respectively. It is interesting to note that even though Japan faces the highest tariff rates imposed by Cambodia, it in turn imposes only about 0.16 percent on Cambodian goods on average. Perhaps, one reason that can explain this is that after World War 2, the Cambodian government declined to demand reparations for the atrocities and war crimes committed by Japan, and since then, the Japanese government has maintained a cooperative and supportive relationship with Cambodia (Chea 2018).

Country	Trade Balance (in millions USD)	Total Import (in millions USD)	Total Export (in millions USD)	Avg Tariff Imposed by Cambodia (%)	Avg Tariff Imposed by Partner Country (%)
China	-9204.9	10445.53	1240.63	8.23	3.42
France	309.62	119.3	428.92	7.11	0
Hong Kong	-582.33	764.33	182	3.45	0
Japan	398.04	774.99	1173.03	15.53	0.16
Malaysia	-396.67	508.52	111.85	7.57	9.45
Singapore	-3137.75	3230.32	92.58	0.94	0
South Korea	-311.65	545.29	233.64	13.49	2.38
Thailand	-3422.9	3832.5	409.6	11.07	2.55
USA	8658.74	312.63	8971.37	12.45	7.95
Vietnam	-3283.51	3967.14	683.63	6.65	0.54

Figure 6. Table showing Cambodia's overall trade data.

Most of Cambodia’s exports are textiles, and for many countries, like France, Japan, South Korea and the United States, textiles make up the largest share of their imports from Cambodia (Kimsovanna 2025). Vegetables (or agricultural products) are also significant, accounting for about a third of all exports to China and Malaysia mainly. Almost half of the exports to Vietnam are plastic or rubber, which is expected given that out of the 13 projects that Vietnamese companies invested in Cambodia, seven are focused on rubber cultivation. In terms of imports coming into Cambodia, the data gets more interesting. For instance, China exports about $3.7 billion worth of textiles to Cambodia despite the fact that Cambodia itself is a textile

manufacturing hub; perhaps this may be due to comparative advantage, where clothes coming from China are cheaper than those made at home. Both Hong Kong and Singapore export mostly stone and glass products to Cambodia, while the United States, Japan and South Korea, all known for their huge competitive automobile industries, primarily export transportations to Cambodia (i.e., cars and motorcycles). Even though Cambodia has recently stepped up in its energy productions thanks to these FDI projects, it still relies heavily on fuel imports from neighboring countries like Thailand and Vietnam, which combined to about $2.12 billion. These trade patterns show how complex and deeply connected Cambodia's economic ties are with both regional and global partners, often following similar trends as its FDI activity.

Challenges and Limitations

There were several challenges when developing this database application. As discussed before, there were some issues with data availability, such as in the case of Taiwan lacking its trade data, as well as cases where foreign firms did not report their true country of origin, making it difficult to verify, leading to many of them to be excluded from the data. This could explain why there is no data on FDI projects in provinces like Kampong Cham, Oddar Meanchey, Pailin and Kep. Gathering and verifying all the data, cleaning and inserting it into Excel and pgAdmin 4 took up a lot of time for me in this project. Designing the UI was time-consuming, too, but at least it was more enjoyable. There was also a risk of going over the 25-hour limit usage on the Shiny app, as I was using the free version. This means that any time anyone opens my database online, including myself, it will count towards the time limit; thankfully, I still had more than a dozen hours left after completing this project. Furthermore, I know I was relying on AI for guidance when it came to the more intricate codes, but I have been learning new tools and functions from them as well.

Conclusion

Ultimately, developing this application has served two main purposes, which were to improve data accessibility by compiling the relevant information into one database, and to help me obtain insights about how FDI activities in Cambodia reflect on its export and import patterns. Even though there were some constraints and limitations as discussed above, I was able to publish my application, host it onto the Shiny app server, and embed it into my Github website.

References

Chea, Vandith. 2018. "Japan Won't Abandon Cambodia." *Khmer Times*. https://www.khmertimeskh.com/504332/japan-wont-abandon-cambodia/.

Council for the Development of Cambodia (CDC). 2024. "Textile and Apparel." *Council for the Development of Cambodia*. https://cdc.gov.kh/home/textile-and-aparel/.

GeoBoundaries. 2023. "Cambodia - Subnational Administrative Boundaries." *Humanitarian Data Exchange*. https://data.humdata.org/dataset/geoboundaries-admin-boundaries-for-cambodia.

International Monetary Fund (IMF). 2025. "Cambodia: 2024 Article IV Consultation—Press Release; Staff Report; and Statement by the Executive Director for Cambodia." *IMF Staff Country Reports*, no. 2025/022. https://www.imf.org/en/News/Articles/2025/01/27/pr-25017-cambodia-imf-executive-board-concludes-2024-article-iv-consultation-with-cambodia.

Open Development Cambodia (ODC). 2023. "Foreign Investment in Cambodia." *Open Development Cambodia*. https://opendevelopmentcambodia.net/profiles/foreign-investment-in-cambodia/.

Pham, Phan Long. 2024. "Cambodia's Funan Techo Canal Project: A Catalog of Worries (Analysis)." *Mongabay*. July 15. https://news.mongabay.com/2024/07/cambodias-funan-techo-canal-project-a-catalog-of-worries-analysis/.

World Bank Group. 2024. "The World Bank in Cambodia: Overview." *World Bank*. https://www.worldbank.org/en/country/cambodia/overview.

World Integrated Trade Solution (WITS). 2022. "Cambodia Trade Summary 2022." *World*

Integrated Trade Solution.

https://wits.worldbank.org/CountryProfile/en/Country/KHM/Year/2022/Summary.