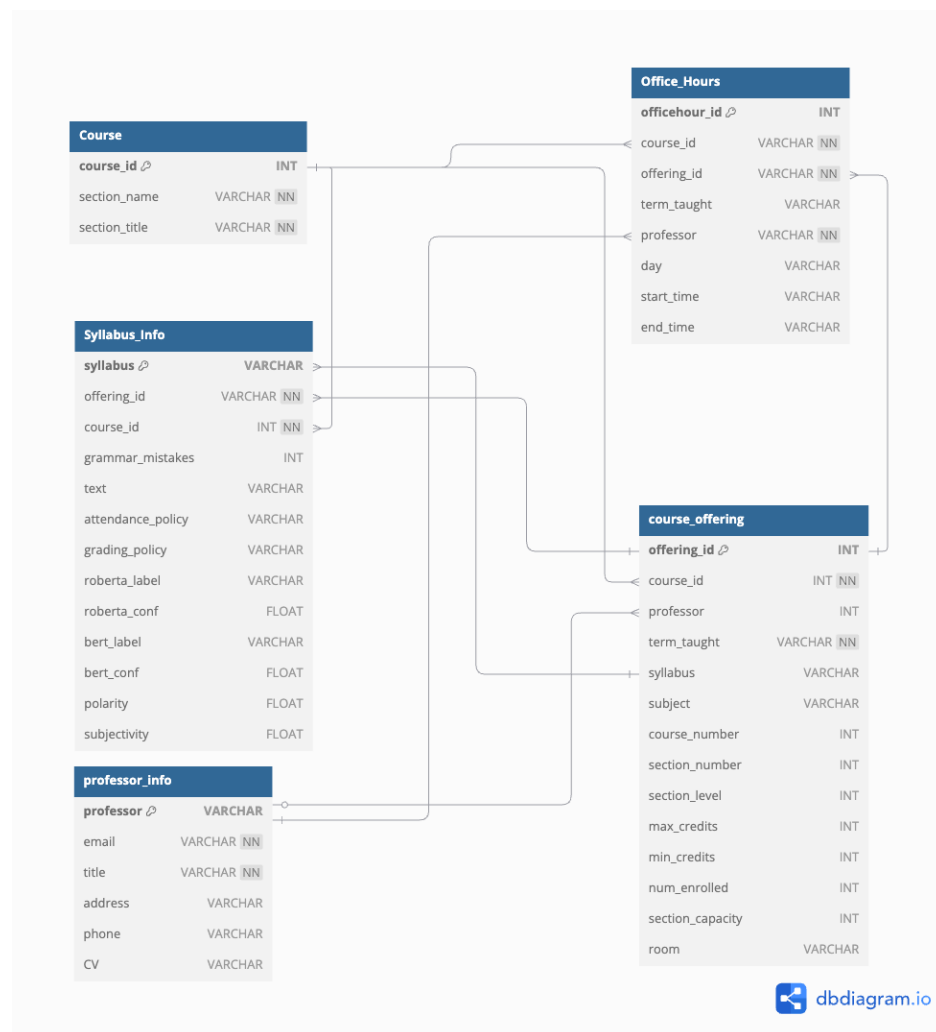


1 Data Sources

All original data was sourced from the Grinnell College website, the Registrar, or the Dean's Office. Information on History classes and professors was extracted from the Grinnell College website. Specifics on how web pages were used are provided below. The Registrar provided additional information on offering in the statistics department, as well as capacity, enrollment, number of credits, and location of the classroom for courses in both the History and Statistics department. The Dean's Office provided pdf files of syllabi which were stored locally and named with the format

DEPARTMENT_coursenumber_coursesection_semester_Professor, such as STA_09_01_SP22_Jonkman.pdf.

2 Database Structure



2.1 course.csv

Contains every distinct course taught in the department.

- `course_id`: Unique identifier for the course. Combination of section name and title.
- `section_name`: Name of the course, i.e. (HIS-100-02)
- `section_title`: Title of the course, i.e. Essays and Writing of History

2.2 course_offering.csv

Contains information on each particular semester that a course was taught in.

- `offering_id`: Combination of the term taught and the course name.
- `course_id`: Unique identifier for the course. Combination of section name and title.
- `professor`: Instructor assigned to the course offering
- `term_taught`: Academic term when the course is offered (e.g., Fall 2024)
- `syllabus`: Link to the course syllabus
- `subject`: Subject of the course
- `section_number`: Number designating a particular section of the course
- `section_level`: Level of the section i.e. 100-level
- `max_credits`: Maximum credits a student can earn from this course
- `min_credits`: Minimum credits a student can earn from this course
- `num_enrolled`: Number of students currently enrolled
- `section_capacity`: Maximum number of students allowed
- `room`: Room where the class is held

2.3 professor_info.csv

Contains information on professor who've taught courses in the department.

- `professor`: Name of the professor.
- `email`: Contact email of the professor
- `title`: Academic or professional title
- `address`: Office location
- `phone`: Contact phone number
- `CV`: Link to the professor's CV

2.4 Table: office_hours.csv

Contains information on the office hours held by professors in this department.

- `officehour_id`: Identifier for a specific office hour entry
- `course_id`: Course linked to the office hour
- `offering_id`: Course offering linked to the office hour
- `term_taught`: Term during which the office hours occur
- `professor`: Professor holding the office hours
- `day`: Day of the week the office hour is held
- `start_time`: Time office hour begins
- `end_time`: Time office hour ends

2.5 syllabus_info.csv

Contains information on the syllabi associated with the course offerings.

- syllabus: URL or reference to the syllabus document
- offering_id: Course offering the syllabus belongs to
- course_id: Course associated with the syllabus
- roberta_label: The final label assigned based on the majority vote of all chunk-level predictions (Positive vs. Negative) by the RoBERTa model.
- roberta_conf: The average confidence score across all RoBERTa chunks.
- bert_Label: The average numeric star rating (like 3.75 stars) computed across all BERT chunks.
- bert_Conf: The average confidence score across all BERT chunk predictions.
- polarity: A float from TextBlob representing sentiment direction, where values closer to +1 indicate positive language and -1 indicate negative sentiment.
- subjectivity: A float from TextBlob indicating how subjective (opinion-based) the text is, with 0 being entirely objective and 1 being entirely subjective.